The Practice of Finitism: Epsilon Calculus and Consistency Proofs in Hilbert's Program

Richard Zach (zach@math.berkeley.edu)
University of California, Berkeley

Second Draft, February 21, 2001 – Comments welcome!

Abstract. After a brief flirtation with logicism in 1917–1920, David Hilbert proposed his own program in the foundations of mathematics in 1920 and developed it, in concert with collaborators such as Paul Bernays and Wilhelm Ackermann, throughout the 1920s. The two technical pillars of the project were the development of axiomatic systems for ever stronger and more comprehensive areas of mathematics and finitistic proofs of consistency of these systems. Early advances in these areas were made by Hilbert (and Bernays) in a series of lecture courses at the University of Göttingen between 1917 and 1923, and notably in Ackermann's dissertation of 1924. The main innovation was the invention of the ε-calculus, on which Hilbert's axiom systems were based, and the development of the ε-substitution method as a basis for consistency proofs. The paper traces the development of the "simultaneous development of logic and mathematics" through the ε-notation and provides an analysis of Ackermann's consistency proofs for primitive recursive arithmetic and for the first comprehensive mathematical system, the latter using the substitution method. It is striking that these proofs use transfinite induction not dissimilar to that used in Gentzen's later consistency proof as well as non-primitive recursive definitions, and that these methods were accepted as finitistic at the time.

1. Introduction

Hilbert first presented his philosophical ideas based on the axiomatic method and consistency proofs in the years 1904 and 1905, following his exchange with Frege on the nature of axiomatic systems and the publication of Russell's Paradox. In the text of Hilbert's address to the International Congress of Mathematicians in Heidelberg, we read:

> Arithmetic is often considered to be part of logic, and the traditional fundamental logical notions are usually presupposed when it is a question of establishing a foundation of arithmetic. If we observe attentively, however, we realize that in the traditional exposition of the laws of logic certain fundamental arithmetic notions are already used, for example, the notion of set and, to some extent, also that of number. Thus we find ourselves turning in a circle, and that is why a partly simultaneous development of the laws of logic and arithmetic is required if paradoxes are to be avoided.[1]

When Hilbert returned to his foundational work with full force in 1917, he seems at first to have been impressed with Russell's and Whiteheads's work in the *Principia*, which succeeded in developing large parts of mathematics

without using sets. By 1920, however, Hilbert returned to his earlier conviction that a reduction of mathematics to logic is not likely to succeed. Instead, he takes Zermelo's axiomatic set theory as a suitable framework for developing mathematics. He localizes the failure of Russell's logicism in its inability to provide the existence results necessary for analysis:

> The axiomatic method used by Zermelo is unimpeachable and indispensable. The question whether the axioms include a contradiction, however, remains open. Furthermore the question poses itself if and in how far this axiom system can be deduced from logic. [. . . T]he attempt to reduce set theory to logic seems promising because sets, which are the objects of Zermelo's axiomatics, are closely related to the predicates of logic. Specifically, sets can be reduced to predicates.
>
> This idea is the starting point for Frege's, Russell's, and Weyl's investigations into the foundations of mathematics.[2]

The logicist project runs into a difficulty when, given a second-order predicate S to which a set of sets is reduced, we want to know that there is a predicate to which the union of the sets reduces. This predicate would be $(\exists P)(P(x) \& S(P))$— x is in the union of the sets in S if there is a set P of which x is a member and which is a member of S.

> We have to ask ourselves, what "there is a predicate P" is supposed to mean. In axiomatic set theory "there is" always refers to a basic domain $\mathfrak{B}$. In logic we could also think of the predicates comprising a domain, but this domain of predicates cannot be seen as something given at the outset, but the predicates must be formed through logical operations, and the rules of construction determine the domain of predicates only afterwards.
>
> From this we see that in the rules of logical construction of predicates reference to the domain of predicates cannot be allowed. For otherwise a *circulus vitiosus* would result.[3]

Here Hilbert is echoing the predicativist worries of Poincaré and Weyl. However, Hilbert rejects Weyl's answer to the problem, viz., restricting mathematics to predicatively acceptable constructions and inferences, as unacceptable in that it amounts to "a return to the prohibition policies of Kronecker." Russell's proposed solution, on the other hand, amounts to giving up the aim of reduction to logic:

> Russell starts with the idea that it suffices to replace the predicate needed for the definition of the union set by one that is extensionally equivalent, and which is not open to the same objections. He is unable, however, to exhibit such a predicate, but sees it as obvious that such a predicate exists. It is in this sense that he postulates the "axiom of reducibility," which states approximately the following: "For each predicate, which is formed by referring (once or multiple times) to the domain of predicates,

> there is an extensionally equivalent predicate, which does not make such reference.
>
> With this, however, Russell returns from constructive logic to the axiomatic standpoint. [. . .]
>
> The aim of reducing set theory, and with it the usual methods of analysis, to logic, has not been achieved today and maybe cannot be achieved at all.[4]

With this, Hilbert rejects the logicist position as failed. At the same time, he rejects the restrictive positions of Brower, Weyl, and Kronecker. The axiomatic method provides a framework which can accommodate the positive contributions of Brouwer and Weyl, without destroying mathematics through a Kroneckerian "politics of prohibitions." For Hilbert, the unfettered progress of mathematics, and science in general, is a prime concern. This is a position that Hilbert had already stressed in his lectures before the 1900 and 1904 International Congresses of Mathematics, and which is again of paramount importance for him with the conversion of Weyl to Brouwer's intuitionism.

Naturally, the greater freedom comes with a price attached: the axiomatic method, in contrast to a foundation based on logical principles alone, does not itself guarantee consistency. Thus, a proof of consistency is needed.

2. Early Consistency Proofs

Ever since his work on geometry in the 1890s, Hilbert had an interest in consistency proofs. The approaches he used prior to the foundational program of the 1920s were almost always relative consistency proofs. Various axiomatic systems, from geometry to physics, were shown to be consistent by giving arithmetical (in a broad sense, including arithmetic of the reals) interpretations for these systems, with one exception—a prototype of a finitistic consistency proof for a weak arithmetical system in Hilbert (1905). This was Hilbert's first attempt at a "direct" consistency proof for arithmetic, i.e., one not based on a reduction to another system, which he had posed as the second of his famous list of problems (Hilbert, 1900).

When Hilbert once again started working on foundational issues following the war, the first order of business was a formulation of logic. This was accomplished in collaboration with Bernays between 1917 and 1920 (see Sieg, 1999 and Zach, 1999), included the establishment of metatheoretical results like completeness, decidability, and consistency for propositional logic in 1917/18, and was followed by ever more nuanced axiom systems for propositional and predicate logic. This first work in purely logical axiomatics was soon extended to include mathematics. Here Hilbert followed his own proposal, made first in 1905,[5] to develop mathematics and logic simultaneously. The extent of this simultaneous development is nowhere clearer than in Hilbert's

lecture course of 1921/22, where the ε-operator is first used as both a logical notion, representing the quantifiers, and an arithmetical notion, representing induction in the form of the least number principle. Hilbert realized then that a consistency proof for all of mathematics is a difficult undertaking, best attempted in stages:

> Considering the great variety of connectives and interdependencies exhibited by arithmetic, it is obvious from the start that we will the not be able to solve the problem of proving consistency in one fell swoop. We will instead first consider the simplest connectives, and then proceed to ever higher operations and inference methods, whereby consistency has to be established for each extension of the system of signs and inference rules, so that these extensions do not endanger the consistency [result] established in the preceding stage.
>
> Another important aspect is that, following our plan for the complete formalization of arithmetic, we have to develop the proper mathematical formalism in connection with the formalism of the logical operations, so that—as I have expressed it—a simultaneous construction of mathematics and logic is executed.[6]

Hilbert had rather clear ideas, once the basic tools both of proof and of formalization were in place, of what the stages should be. In an addendum to the lecture course on *Grundlagen der Mathematik*, taught by Hilbert and Bernays in 1922–23,[7] he outlined them. The first stage had already been accomplished: Hilbert gave consistency proofs for calculi of propositional logic in his 1917/18 lectures. Stage II consist in the elementary calculus of free variables, plus equality axioms and axioms for successor and predecessor. The axioms are:

1.	$A \to B \to A$	2.	$(A \to A \to B) \to A \to B$
3.	$(A \to B \to C) \to (B \to A \to C)$	4.	$(B \to C) \to (A \to B) \to A \to C$
5.	$A\&B \to A$	6.	$A\&B \to B$
7.	$A \to B \to A\&B$	8.	$A \to A \vee B$
9.	$B \to A \vee B$	10.	$(A \to C) \to (B \to C) \to A \vee B \to C$
11.	$A \to \overline{A} \to B$	12.	$(A \to B) \to (\overline{A} \to B) \to B$
13.	$a = a$	14.	$a = b \to A(a) \to A(b)$
15.	$a + 1 \neq 0$	16.	$\delta(a+1) = a$[8]

In Hilbert's systems, Latin letters are variables; in particular, a, b, c, ..., are individual variables and A, B, C, ..., are formula variables. The rules of inference are modus ponens and substitution for individual and formula variables.

Hilbert envisaged his foundational project as a stepwise "simultaneous development of logic and mathematics," in which axiomatic systems for logic, arithmetic, analysis, and finally set theory would be developed. Each stage would require a proof of consistency before the next stage is developed. In a

handwritten supplement to the typescript of the 1922–23 lecture notes on the foundations of arithmetic, Hilbert presents a rough overview of what these steps might be:

> **Outline**. Stage II was elementary calculation, axioms 1–16.
> Stage III. Now elementary number theory
> Schema for definition of functions by recursion and modus ponens
> will add the schema of induction to modus ponens
> even if this coincides in substance with the results of intuitively obtained number theory, we are now dealing with formulas, e.g, $a+b=b+a$.
> Stage IIII. Transfinite inferences and parts of analysis
> Stage V. Higher-order variables and set theory. Axiom of choice.
> Stage VI. Numbers of the 2nd number class, full transfinite induction. Higher types. Continuum problem, transfinite induction for numbers in the 2nd number class.
> Stage VII. (1) Replacement of infinitely many definitional schemata by one axiom. (2) Analysis and set theory. At level 4, again the full theorem of the least upper bound.
> Stage VIII. Formalization of well ordering.[9]

2.1. The propositional calculus and the calculus of elementary computation

Step I had been achieved in 1917–18. Already in the lectures from the Winter term 1917/18, Hilbert and Bernays had proved that the propositional calculus is consistent. This was done first by providing an arithmetical interpretation, where they stressed that only finitely many numbers had to be used as "values" (0 and 1). The proof is essentially a modern proof of the soundness of propositional logic: A truth value semantics is introduced by associating which each formula of the propositional calculus a truth function mapping tuples of 0 and 1 (the values of the propositional variables) to 0 or 1 (the truth value of the formula under the corresponing valuation). A formula is called *correct* if it corresponds to a truth function which always takes the value 1. It is then showed that the axioms are correct, and that modus ponens preserves correctness. So every formula derivable in the propositional calculus is correct. Since A and $\overline{A}$ cannot both be correct, they cannot both be derivable, and so the propositional calculus is consistent.

It was very important for Hilbert that the model for the propositional calculus thus provided by $\{0,1\}$ was finite. As such, its existence, and the admissibility of the consistency proof was beyond question. This lead him to consider the consistency proof for the propositional calculus to be the prime example for for a consistency proof *by exhibition* in his 1921/22 lectures on the foundations of mathematics. The consistency problem in the form of a demand for a consistency proof for an axiomatic system which neither

proceeds by exhibiting a model, nor by reducing consistency of a system to the consistency of another, but by providing a metamathematical proof that no derivation of a contradiction is possible, is first formulated in lectures in the Summer term of 1920. Here we find a first formulation of an arithmetical system and a proof of consistency. The system consists of the axioms

$$\begin{aligned} 1 &= 1 \\ (a=b) &\rightarrow (a+1=b+1) \\ (a+1=b+1) &\rightarrow (a=b) \\ (a=b) &\rightarrow ((a=c) \rightarrow (b=c)). \end{aligned}$$

The notes contain a proof that these four axioms, together with modus ponens, do not allow the derivation of the formula

$$a+1=1.$$

The proof itself is not too interesting, and I will not reproduce it here.[10] The system considered is quite weak. It does not even contain all of propositional logic: negation only appears as inequality, and only formulas with at most two '$\rightarrow$' signs are derivable. Not even $a=a$ is derivable. It is here, nevertheless, that we find the first statement of the most important ingredient of Hilbert's project, namely, proof theory:

> Thus we are led to make the proofs themselves the object of our investigation; we are urged towards a *proof theory*, which operates with the proofs themselves as objects.
>
> For the way of thinking of ordinary number theory the numbers are then objectively exhibitable, and the proofs about the numbers already belong to the area of thought. In our study, the proof itself is something which can be exhibited, and by thinking about the proof we arrive at the solution of our problem.
>
> Just as the physicist examines his apparatus, the astronomer his position, just as the philosopher engages in critique of reason, so the mathematician needs his proof theory, in order to secure each mathematical theorem by proof critique.[11]

This project is developed in earnest in two more lecture courses in 1921–22 and 1922–23. These lectures are important in two respects. First, it is here that the axiomatic systems whose consistency is to be proven are developed. This is of particular interest for an understanding of the relationship of Hilbert to Russell's project in the *Principia* and the influence of Russell's work both on Hilbert's philosophy and on the development of axiomatic systems for mathematics.[12] Sieg (1999) has argued that, in fact, Hilbert was a logicist for a brief period around the time of his paper "Axiomatic Thought" (Hilbert, 1918a). However, as noted in Section 1, Hilbert soon became critical of Russell's

type theory, in particular of the axiom of reducibility. Instead of taking the system of *Principia* as the adequate formalization of mathematics the consistency of which was to be shown, Hilbert proposed a new system. The guiding principle of this system was the "simultaneous development of logic and mathematics"—as opposed to a development of mathematics out of logic—which he had already proposed in Hilbert (1905, 176). The cornerstone of this development is the ε-calculus. The second major contribution of the 1921–22 and 1922–23 lectures are the consistency proofs themselves, including the *Hilbertsche Ansatz* for the ε-substitution method, which were the direct precursors to Ackermann's dissertation of 1924.

In contrast to the first systems of 1920, here Hilbert uses a system based on full propositional logic with axioms for equality, i.e., the elementary calculus of free variables:

I. Logical axioms
a) Axioms of consequence
1) $A \to B \to A$
2) $(A \to A \to B) \to A \to B$
3) $(A \to B \to C) \to B \to A \to C$
4) $(B \to C) \to (A \to B) \to A \to C$

b) Axioms of negation
5) $A \to \overline{A} \to B$
6) $(A \to B) \to (\overline{A} \to B) \to B$
II. Arithmetical axioms
a) Axioms of equality
7) $a = a$
8) $a = b \to Aa \to Ab$
b) Axioms of number
9) $a + 1 \neq 0$
10) $\delta(a+1) = a$[13]

Here, '$+1$' is a unary function symbol. The rules of inference are substitution (for individual and formula variables) and modus ponens.

Hilbert's idea for how a finitistic consistency proof should be carried out is first presented here. The idea is this: suppose a proof of a contradiction is available. We may assume that the end formula of this proof is $0 \neq 0$.

1. *Resolution into proof threads.* First, we observe that by duplicating part of the proof and leaving out steps, we can transform the derivation to one where each formula (except the end formula) is used exactly once as the premise of an inference. Hence, the proof is in tree form.

2. *Elimination of variables.* We transform the proof so that it contains no free variables. This is accomplished by proceeding backwards from the

end formula: The end formula contains no free variables. If a formula is the conclusion of a substitution rule, the inference is removed. If a formula is the conclusion of modus ponens it is of the form

$$\frac{\mathfrak{A} \quad \mathfrak{A} \to \mathfrak{B}}{\mathfrak{B}'}$$

where $\mathfrak{B}'$ results from $\mathfrak{B}$ by substituting terms for free variables. If these variables also occur in $\mathfrak{A}$, we substitute the same terms for them. Variables in $\mathfrak{A}$ which do not occur in $\mathfrak{B}$ are replaced with 0. This yields a formula $\mathfrak{A}'$ not containing variables.[14] The inference is replaced by

$$\frac{\mathfrak{A}' \quad \mathfrak{A}' \to \mathfrak{B}'}{\mathfrak{B}'}$$

3. *Reduction of functionals.* The remaining derivation contains a number of terms (*functionals* in Hilbert's parlance) which now have to be reduced to numerical terms (i.e., standard numerals of the form $(\ldots(0+1)+\cdots)+1)$. In this case, this is done easily by rewriting innermost subterms of the form $\delta(0)$ by 0 and $\delta(\mathfrak{n}+1)$ by $\mathfrak{n}$. In later stages, the set of terms is extended by function symbols introduced by recursion, and the reduction of functionals there proceeds by calculating the function for given numerical arguments according to the recursive definition. This will be discussed in the next section.

In order to establish the consistency of the axiom system, Hilbert suggests, we have to find a decidable (*konkret feststellbar*) property of formulas so that every formula in a derivation which has been transformed using the above steps has the property, and the formula $0 \neq 0$ lacks it. The property Hilbert proposes to use is *correctness*. This is not to be understood as truth in a model. The formulas still occurring in the derivation after the transformation are all Boolean combinations of equations between numerals. An equation between numerals $\mathfrak{n} = \mathfrak{m}$ is *correct* if $\mathfrak{n}$ and $\mathfrak{m}$ are syntactically equal, and the negation of an equality is correct of $\mathfrak{m}$ and $\mathfrak{n}$ are not syntactically equal.

> If we call a formula which does not contain variables or functionals other than numerals an "*explicit [numerical] formula*", then we can express the result obtained thus: Every provable explicit [numerical] formula is end formula of a proof all the formulas of which are explicit formulas.
>
> This would have to hold in particular of the formula $0 \neq 0$, if it were provable. The required proof of consistency is thus completed if we show that there can be no proof of the formula which consists of only explicit formulas.
>
> To see that this is impossible it suffices to find a concretely determinable [*konkret feststellbar*] property, which first of all holds of all explicit

> formulas which result from an axiom by substitution, which furthermore transfers from premises to end formula in an inference, which however does not apply to the formula $0 \neq 0$.[15]

Hilbert now defines the notion of a (conjunctive) normal form and gives a procedure to transform a formula into such a normal form. He then provides the wanted property:

> With the help of the notion of a normal form we are now in a position to exhipit a property which distinguishes the formula $0 \neq 0$ from the provable explicit formulas.
>
> We divide the explicit formulas into "*correct*" and "*incorrect*." The explicit atomic formulas are equations with *numerals* on either side [of the equality symbol]. We call such an *equation correct*, if the numerals on either side *coincide*, otherwise we call it *incorrect*. We call an *inequality* with numerals on either side *correct* if the two numerals are *different*, otherwise we call it *incorrect*.
>
> In the normal form of an arbitrary explicit formula, each disjunct has the form of an equation or an inequality with numerals on either side.
>
> We now call a *general explicit formula correct* if in the corresponding normal form each disjunction which occurs as a conjunct (or which constitutes the normal form) contains a correct equation or a correct inequality. Otherwise we call the formula *incorrect*. [. . .]
>
> According to this definition, the question of whether an explicit formula is correct or incorrect is *concretely decidable* in every case. Thus the "tertium non datur" holds here. . . [16]

This use in the 1921–22 lectures of the conjunctive normal form of a propositional formula to define correctness of Boolean combinations of equalities between numerals goes back to the 1917–18 lecture notes,[17] where transformation into conjunctive normal form and testing whether each conjunct contains both A and $\overline{A}$ was proposed as a test for propositional validity. Similarly, here a formula is *correct* if each conjunct in its conjunctive normal form contains a correct equation or a correct inequality.[18] In the 1922–23 lectures, the definition involving conjunctive normal forms is replaced by the usual inductive definition of propositional truth and falsehood by truth tables (Hilbert and Bernays 1923a, 21). Armed with the definition of correct formula, Hilbert can prove that the derivation resulting from a proof by transforming it according to (1)–(3) above contains only correct formulas. Since $0 \neq 0$ is plainly not correct, there can be no proof of $0 \neq 0$ in the system consisting of axioms (1)–(10). The proof is a standard induction on the length of the derivation: the formulas resulting from the axioms by elimination of variables and reduction of functionals are all correct, and modus ponens preserves correctness.[19]

2.2. Elementary number theory with recursion and induction rule

The system of stage III consists of the basic system of the elementary calculus of free variables and the successor function, extended by the schema of defining functions by primitive recursion and the induction rule.[20] A primitive recursive definition is a pair of axioms of the form

$$\begin{aligned} \varphi(0, b_1, \ldots, b_n) &= \mathfrak{a}(b_1, \ldots, b_n) \\ \varphi(a+1, b_1, \ldots, b_n) &= \mathfrak{b}(a, \varphi(a), b_1, \ldots, b_n) \end{aligned}$$

where $\mathfrak{a}(b_1, \ldots, b_n)$ contains only the variables $b_1, \ldots, b_n$, and $\mathfrak{b}(a, c, b_1, \ldots, b_n)$ contains only the variables a, c, b_1, $\ldots$, b_n. Neither contains the function symbol φ or any function symbols which have not yet been defined.

The introduction of primitive recursive definitions and the induction rule serves, first of all, the purpose of expressivity. Surely any decent axiom system for arithmetic must provide the means of expressing basic number-theoretic states of affairs, and this includes addition, subtraction, multiplication, division, greatest common divisor, etc. The general schema of primitive recursion is already mentioned in the Kneser notes for 1921–22 (Hilbert, 1922a, Heft II, 29), and is discussed in some detail in the notes for the lectures of the following year (Hilbert and Bernays, 1923a, 26–30).

It may be interesting to note that in the 1922–23 lectures, there are no axioms for addition or multiplication given before the general schema for recursive definition. This suggests a change in emphasis during 1922, when Hilbert realized the importance of primitive recursion as an arithmetical concept formation. He later continued to develop the notion, hoping to capture all number theoretic functions using an extended notion of primitive recursion and to solve the continuum problem with it. This can be seen from the attempt at a proof of the continuum hypothesis in (1926), and Ackermann's paper on "Hilbert's construction of the reals" (1928b), which deals with hierarchies of recursive functions. The general outlook in this regard is also markedly different from Skolem's 1923, which is usually credited with the definition of primitive recursive arithmetic.[21]

Hilbert would be remiss if he would not be including induction in his arithmetical axiom systems. As he already indicates in the 1921–22 lectures, however, the induction principle cannot be formulated as an axiom without the help of quantifiers.

> We are still completely missing the axiom of complete induction. One might think it would be
>
> $$\{Z(a) \to (A(a) \to A(a+1))\} \to \{A(1) \to (Z(b) \to A(b))\}$$

> That is not it, for take $a = 1$. The hypothesis must hold for *all a*. We have, however, no means to bring the *all* into the hypothesis. Our formalism does not yet suffice to write down the axiom of induction.
>
> But as a schema we can: We extend our methods of proof by the following schema.
>
> $$\frac{\mathfrak{K}(1) \quad \mathfrak{K}(a) \to \mathfrak{K}(a+1)}{Z(a) \to \mathfrak{K}(a)}$$
>
> Now it makes sense to ask whether this schema can lead to a contradiction.[22]

The induction schema is thus necessary in the formulation of the elementary calculus only because quantifiers are not yet available. Subsequently, induction will be subsumed in the ε-calculus.

The consistency proof for stage II is extended to cover also the induction schema and primitive recursive definitions. Both are only sketched: Step (3), reduction of functionals, is extended to cover terms containing primitive recursive functions by recursively computing the value of the innermost term containing only numerals. Both in the 1921–22 and the 1922–23 sets of notes by Kneser, roughly a paragraph is devoted to these cases (the official sets of notes for both lectures do not contain the respective passages).

> How de we proceed for recursions? Suppose a $\varphi(\mathfrak{z})$ occurs. Either [$\mathfrak{z}$ is] 0, then we replace it by $\mathfrak{a}$. Or [it is of the form] $\varphi(\mathfrak{z}+\mathfrak{1})$: [replace it with] $\mathfrak{b}(\mathfrak{z}, \varphi(\mathfrak{z}))$. Claim: These substitutions eventually come to an end, if we replace innermost occurrences first.[23]

The claim is not proved, and there is no argument that the process terminates even for terms containing several different, nested primitive recursively defined function symbols.

For the induction schema, Hilbert hints at how the consistency proof must be extended. Combining elimination of variables and reduction of functionals we are to proceed upwards in the proof as before until we arrive at an instance of the induction schema:

$$\frac{\mathfrak{K}(1) \quad \mathfrak{K}(a) \to \mathfrak{K}(a+1)}{Z(\mathfrak{z}) \to \mathfrak{K}'(\mathfrak{z})}$$

By copying the proof ending in the right premise, substituting numerals 1, ..., $\mathfrak{y}$ (where $\mathfrak{z} = \mathfrak{y} + 1$) for a and applying the appropriate substitutions to the other variables in $\mathfrak{K}$ we obtain a proof of $Z(a) \to \mathfrak{K}'(\mathfrak{z})$ without the last application of the induction schema.

With the introduction of the ε-calculus, the induction rule is of only minor importance, and its consistency is never proved in detail until Hilbert and Bernays (1934, 298–99).

2.3. THE ε-CALCULUS AND THE AXIOMATIZATION OF MATHEMATICS

In the spirit of the "simultaneous development of logic and mathematics," Hilbert takes the next step in the axiomatization of arithmetic by employing a principle taken from Zermelo's axiomatization of set theory: the axiom of choice. Hilbert and Bernays had dealt in detail with quantifiers in lectures in 1917–18 and 1920, but they do not directly play a significant role in the axiom systems Hilbert develops for mathematics. Rather, the first- and higher-order calculi for which consistency proofs are proposed, are based instead on choice functions. The first presentation of these ideas can be found in the 1921–22 lecture notes by Kneser (the official notes do not contain these passages). The motivation is that in order to deal with analysis, one has to allow definitions of functions which are not finitary. These concept formations, necessary for the development of mathematics free from intuitionist restrictions, include definition of functions from undecidable properties, by unbounded search, and choice.

> Not finitely (recursively) defined is, e.g., $\varphi(a) = 0$ if there is a b so that $a^5 + ab^3 + 7$ is prime, and $= 1$ otherwise. But only with these numbers and functions the real mathematical interest begins, since the solvability in finitely many steps is not foreseeable. We have the conviction, that such questions, e.g., the value of $\varphi(a)$, are solvable, i.e., that $\varphi(a)$ is also finitely definable. We cannot wait on this, however, we must allow such definitions for otherwise we would restrict the free practice of science. We also need the concept of a function of functions.[24]

The concepts which Hilbert apparently takes to be fundamental for this project are the principle of the excluded middle and the axiom of choice, in the form of second-order functions τ and α. The axioms for these functions are

1. $\tau(f) = 0 \to (Z(a) \to f(a) = 1)$
2. $\tau(f) \neq 0 \to Z(\alpha(f))$
3. $\tau(f) \neq 0 \to f(\alpha(f)) \neq 1$
4. $\tau(f) \neq 0 \to \tau(f) = 1$

The intended interpretation is: $\tau(f) = 0$ if f is always 1 and $= 1$ if one can choose an $\alpha(f)$ so that $f(\alpha(f)) \neq 1$.

The introduction of τ and α allows Hilbert to replace universal and existential quantifiers, and also provides the basis for proofs of the axiom of induction and the least upper bound principle. Furthermore, Hilbert claims, the consistency of the resulting system can be seen in the same way used to establish the consistency of stage III (primitive recursive arithmetic). From a proof of a numerical formula using τ's and α's, these terms can be eliminated by

finding numerical substitutions which turn the resulting formulas into correct numerical formulas.

These proofs are sketched in the last part of the 1921–22 lecture notes by Kneser (Hilbert, 1922a).[25] In particular, the consistency proof contains the entire idea of the *Hilbertsche Ansatz*, the ε-substitution method:

> First we show that we can eliminate all variables, since here also only free variables occur. We look for the innermost τ and α. Below these there are only finitely defined [primitive recursive] functions φ, φ'. Some of these functions can be substituted for f in the axioms in the course of the proof. 1: $\tau(\varphi) = 0 \rightarrow (Z(\mathfrak{a}) \rightarrow \varphi(\mathfrak{a}) = 1)$, where $\mathfrak{a}$ is a functional. If this is *not* used, we set all $\alpha(\varphi)$ and $\tau(\varphi)$ equal to zero. Otherwise we reduce $\mathfrak{a}$ and $\varphi(\mathfrak{a})$ and check whether $Z(\mathfrak{a}) \rightarrow \varphi(\mathfrak{a}) = 1$ is correct everywhere it occurs. If it is correct, we set $\tau[\varphi] = 0$, $\alpha[\varphi] = 0$. If it is incorrect, i.e., if $\mathfrak{a} = \mathfrak{z}$, $\varphi(\mathfrak{z}) \neq 1$, we let $\tau(\varphi) = 1$, $\alpha(\varphi) = \mathfrak{z}$. After this replacement, the proof remains a proof. The formulas which take the place of the axioms are correct.
>
> (The idea is: if a proof is given, we can extract an argument from it for which $\varphi = 1$.) In this way we eliminate the τ and α and applications of [axioms] (1)–(4) and obtain a proof of $1 \neq 1$ from I–V and correct formulas, i.e., from I–V,
>
> $$\begin{aligned} \tau(f,b) = 0 \ &\rightarrow\ \{Z(a) \rightarrow f(a,b) = 1\} \\ \tau(f,b) \neq 0 \ &\rightarrow\ Z(f(\alpha,b)) \\ \tau(f,b) \neq 0 \ &\rightarrow\ f(\alpha(f,b),b) \neq 1 \\ \tau(f,b) \neq 0 \ &\rightarrow\ \tau(f,b) = 1 \end{aligned}$$ [26]

Although not formulated as precisely as subsequent presentations, all the ingredients of Hilbert's ε-substitution method are here. The only changes that are made en route to the final presentation of Hilbert's sketch of the case for one ε and Ackermann's are mostly notational. In (Hilbert, 1923), a talk given in September 1922, the two functions τ and α are merged to one function (also denoted τ), which in addition provides the *least* witness for $\tau(f(a)) \neq 1$. There the τ function is also applied directly to formulas. In fact, $\tau_a A(a)$ is the primary notion, denoting the least witness a for which $A(a)$ is false; $\tau(f)$ is defined as $\tau_a(f(a) = 0)$. Interestingly enough, the sketch given there for the substitution method is for the τ-function for functions, not formulas, just as it was in the 1921–22 lectures.

The most elaborate discussion of the ε-calculus can be found in Hilbert's and Bernays's course of 1922–23. Here, again, the motivation for the ε-function is Zermelo's axiom of choice:

> What are we missing?

1. As far as logic is concerned: we have had the propositional calculus extended by free variables, i.e., variables for which arbitrary functionals may be substituted. Operating with "all" and "there is" is still missing.
2. We have added the induction schema, but without consistency proof and also on a provisional basis, with the intention of removing it.
3. So far only the arithmetical axioms which refer to whole numbers. The above shortcomings prevent us from building up analysis (limit concept, irrational number).

These 3 points already give us a plan and goals for the following.

We turn to (1). It is clear that a logic without "all'—"there is" would be incomplete, I only recall how the application of these concepts and of the so-called transfinite inferences has brought about major problems. We have not yet addressed the question of the applicability of these concepts to infinite totalities. Now we could proceed as we did with the propositional calculus: Formulate a few simple [principles] as axioms, from which all others follow. Then the consistency proof would have to be carried out—according to our general program: with the attitude that a proof is a figure given to us. Significant obstacles to the consistency proof because of the bound variables. The deeper investigation, however, shows that the real core of the problem lies at a different point, to which one usually only pays attention later, and which also has only been noticed in the literature of late.[27]

At the corresponding place in the Kneser *Mitschrift*, Hilbert continues:

> [This core lies] in Zermelo's *axiom of choice*. [...] The objections [of Brouwer and Weyl] are directed against the choice principle. But they should likewise be directed against "all" and "there is", which are based on the same basic idea.
>
> We want to extend the axiom of choice. To each proposition with a variable $A(a)$ we assign an object for which the proposition holds only if is holds in general. So, a counterexample, if one exists.
>
> $\varepsilon(A)$, an individual logical function. [...] ε satisfies the *transfinite axiom*:
>
> (16) $A(\varepsilon A) \to Aa$
>
> e.g., Aa means: a is corrupt. εA is Aristides.[28]

Hilbert goes on to show how quantifiers can be replaced by ε-terms. The corresponding definitional axioms are already included in Hilbert (1923), i.e., $A(\varepsilon A) \equiv (a)A(a)$ and $A(\varepsilon\overline{A}) \equiv (\exists a)A(a)$. Next, Hilbert outlines a derivation of the induction axioms using the ε-axioms. For this, it is necessary to require that the choice function takes the minimal value, which is expressed by the additional axiom

$$\varepsilon A \neq 0 \to A(\delta(\varepsilon A)).$$

With this addition, Hilbert combined the κ function of (Hilbert, 1922c) and the μ function of (1923) with the ε function. Both κ ("k" for *Kleinstes*, least) and μ had been introduced there as functions of functions giving the least value for which the function differs from 0. In Hilbert (1923, 161–162), Hilbert indicates that the axiom of induction can be derived using the μ function, and credits this to Dedekind (1888).

The third issue Hilbert addresses is that of dealing with real numbers, and extending the calculus to analysis. A first step can be carried out at stage IV by considering a real number as a function defining an infinite binary expansion. A sequence of reals can then be given by a function with two arguments. Already in Hilbert (1923) we find a sketch of the proof of the least upper bound principle for such a sequence of reals, using the π function:

$$\pi A(a) = \begin{cases} 0 & \text{if } (a)A(a) \\ 1 & \text{otherwise} \end{cases}$$

The general case of sets of reals needs function variables and second-order ε and π. These are briefly introduced as $\varepsilon_f A$ with the axioms

$$\begin{aligned} A\varepsilon_f A &\rightarrow Af \\ (f)Af &\rightarrow \pi_f Af = 0 \\ \overline{(f)}Af &\rightarrow \pi_f Af = 1 \end{aligned}$$

The last two lectures transcribed in (Hilbert and Bernays, 1923a) are devoted to a sketch of the ε substitution method. The proof is adapted from Hilbert (1923), replacing εf with εA, also deals with π, and covers the induction axiom in its form for the ε-calculus.[29] During the last lecture, Bernays also extends the proof to second-order ε's.

> If we have a *function variable*:
>
> $$A\varepsilon_f Af \rightarrow Af$$
>
> [...] Suppose ε only occurs with $\mathfrak{A}$ (e.g., $f0 = 0$, $ff0 = 0$). How will we eliminate the function variables? We simply replace fc by c. This does not apply to *bound* variables. For those we take some fixed function, e.g., δ and carry out the reduction with it. Then we are left with, e.g., $\mathfrak{A}\delta \rightarrow \mathfrak{A}\varphi$. This, when reduced, is either correct or incorrect. In the latter case, $\mathfrak{A}\varphi$ is incorrect. Then we substitute φ everywhere for $\varepsilon_f \mathfrak{A} f$. Then we end up with $\mathfrak{A}\varphi \rightarrow \mathfrak{A}\psi$. That is certainly correct, since $\mathfrak{A}\varphi$ is incorrect.[30]

The last development regarding the ε-calculus before Ackermann's dissertation is the switch to the dual notation. Both (Hilbert, 1923) and (Hilbert and Bernays, 1923a) use εA as denoting a counterexample for A, whereas at least from Ackermann's dissertation onwards, εA denotes a witness. Correspondingly, Ackermann uses the dual axiom $A(a) \rightarrow A(\varepsilon_a A(a))$. Although it is relatively

clear that the supplement to (Hilbert and Bernays, 1923a)—24 sheets in Hilbert's hand—are Hilbert's notes based on which he and partly Bernays presented the 1922–23 lectures, parts of it seem to have been altered or written after the conclusion of the course. Sheets 12–14 contain a sketch of the proof of the axiom of induction from the standard, dual ε axioms; the same proof for the original axioms can be found on sheets 8–11.

This concludes the development of mathematical systems using the ε-calculus and consistency proofs for them presented by Hilbert himself. We now turn to the more advanced and detailed treatment in Wilhelm Ackermann's (1924b) dissertation.

3. Ackermann's Dissertation

Wilhelm Ackermann was born in 1896 in Westphalia. He studied mathematics, physics, and philosophy in Göttingen between 1914 and 1924, serving in the army in World War I from 1915–1919. He completed his studies in 1924 with a dissertation, written under Hilbert, entitled "Begründung des 'tertium non datur' mittels der Hilbertschen Theorie der Widerspruchsfreiheit" (Ackermann, 1924a; 1924b), the first major contribution to proof theory and Hilbert's Program. In 1927 he decided for a career as a high school teacher rather than a career in academia, but remained scientifically active. His major contributions to logic include the function which carries his name—an example of a recursive but not primitive recursive function (Ackermann, 1928b), the consistency proof for arithmetic using the ε-substitution method (Ackermann, 1940), and his work on the decision problem (Ackermann, 1928a; 1954). He served as co-author, with Hilbert, of the influential logic textbook *Grundzüge der theoretischen Logik* (Hilbert and Ackermann, 1928). He died in 1962.[31]

Ackermann's 1924 dissertation is of particular interest since it is the first non-trivial example of what Hilbert considered to be a finitistic consistency proof. Von Neumann's paper of 1927 does not entirely fit into the tradition of the Hilbert school, and we have no evidence of the extent of Hilbert's involvement in its writing. Later consistency proofs, in particular those by Gentzen and Kalmár, were written after Gödel's incompleteness results were already well-known and their implications understood by proof theorists. Ackermann's work, on the other hand, arose entirely out of Hilbert's research project, and there is ample evidence that Hilbert was aware of the range and details of the proof. Hilbert was Ackermann's dissertation advisor, approved the thesis, was editor of *Mathematische Annalen,* where the thesis was published, and corresponded with Ackermann on corrections and extensions of the result. Ackermann was also in close contact with Paul Bernays, Hilbert's assistant and close collaborator in foundational matters. Ackermann spent the first half of 1925 in Cambridge, supported by a fellowship from the International Ed-

ucation Board (founded by John D. Rockefeller, Jr., in 1923). In his letter of recommendation for Ackermann, Hilbert writes:

> In his thesis "Foundation of the 'tertium non datur' using Hilbert's theory of consistency," Ackermann has shown in the most general case that the use of the words "all" and "there is," of the "tertium non datur," is free from contradiction. The proof uses exclusively primitive and finite inference methods. Everything is demonstratded, as it were, directly on the mathematical formalism.
>
> Ackermann has here surmounted considerable mathematical difficulties and solved a problem which is of first importance to the modern efforts directed at providing a new foundation for mathematics.[32]

Further testimony of Hilbert's high esteem for Ackermann can be found in the draft of a letter to Russell asking for a letter of support to the International Education Board, where he writes that "Ackermann has taken my classes on foundations of mathematics in recent semesters and is currently one of the best masters of the theory which I have developed here."[33]

Ackermann's work provides insight into two important issues relating to Hilbert's program as it concerns finitistic consistency proofs. First, it provides historical insight into the aims and development of Hilbert's Program: The first part of the program called for an axiomatization of mathematics. These axiomatizations were then the objects of metamathematical investigations: the aim was to find finitistic consistency proofs for them. Which areas of mathematics were supposed to be covered by the consistency proofs, how were they axiomatized, what is the strength of the systems so axiomatized? We have already seen what Hilbert's roadmap for the project of axiomatization was. Ackermann's dissertation provides the earliest example of a formal system stronger than elementary arithmetic. The second aim, the metamathematical investigation of the formal systems obtained, also poses historical questions: When did Ackermann, and other collaborators of Hilbert (in particular, Bernays and von Neumann) achieve the results they sought? Was Ackermann's proof correct, and if not, what parts of it can be made to work?

The other information we can extract from an analysis of Ackermann's work is what methods were used or presupposed in the consistency proofs that were given, and thus, what methods were sanctioned by Hilbert himself as falling under the finitist standpoint. Such an analysis of the methods used are of a deeper, conceptual interest. There is a fundamental division between Hilbert's philosophical remarks on finitism on the one hand, and the professed goals of the program on the other. In these comments, rather little is said about the concept formations and proof methods that a finitist, according to Hilbert, is permitted to use. In fact, most of Hilbert's remarks deal with the objects of finitism, and leave the finitistically admissible forms of definition and proof to the side. These, however, are the questions at issue in contemporary conceptual analyses of finitism. Hilbert's relative silence on the matter

is responsible for the widespread—and largely correct—opinion that Hilbert was too vague on the question of what constitutes finitism to unequivocally define the notion, and therefore later commentators have had enough leeway to disagree widely on the strength of the finitist standpoint while still claiming to have explicated Hilbert's own concept.

3.1. SECOND-ORDER PRIMITIVE RECURSIVE ARITHMETIC

In Ackermann (1924b), the system of stage III is extended by second-order variables for functions. The schema of recursive definition is then extended to include terms containing such variables. In the following outline, I shall follow Ackermann and adopt the notation of subscripting function symbols and terms by variables to indicate that these variables do not occur freely but rather as placeholders for functions. For instance, $\mathfrak{a}_a(f(a))$ indicates that the term $\mathfrak{a}$ does not contain the variable a free, but rather that the function $f(a)$ appears as a functional argument, i.e., that the term is of the form $\mathfrak{a}(\lambda a. f(a))$. The schema of second-order primitive recursion is the following:

$$\begin{aligned}
\varphi_{\vec{b}_i}(0, \vec{f}(\vec{b}_i), \vec{c}) &= \mathfrak{a}_{\vec{b}_i}(\vec{f}(\vec{b}_i), \vec{c}) \\
\varphi_{\vec{b}_i}(a+1, \vec{f}(\vec{b}_i), \vec{c}) &= \mathfrak{b}_{\vec{b}_i}(a, \varphi_{\vec{d}_i}(a, \vec{f}(\vec{d}_i), \vec{c}), \vec{f}(\vec{b}_i))
\end{aligned}$$

To clarify the subscript notation, compare this with the schema of second-order primitive recursion using λ-abstraction notation:

$$\begin{aligned}
\varphi(0, \lambda\vec{b}_i.\vec{f}(\vec{b}_i), \vec{c}) &= \mathfrak{a}(\lambda\vec{b}_i.\vec{f}(\vec{b}_i), \vec{c}) \\
\varphi(a+1, \lambda\vec{b}_i.\vec{f}(\vec{b}_i), \vec{c}) &= \mathfrak{b}(a, \varphi(a, \lambda\vec{d}_i.\vec{f}(\vec{d}_i), \vec{c}), \lambda\vec{b}_i.\vec{f}(\vec{b}_i))
\end{aligned}$$

Using this schema, it is possible to define the Ackermann function. This was already pointed out in Hilbert (1926), although it was not until Ackermann (1928b) that it was shown that the function so defined cannot be defined by primitive recursion without function variables. Ackermann (1928b) defines the function as follows. First it is observed that the iteration function

$$\rho_c(a, f(c), b) = \underbrace{f(\ldots f(f}_{a\ f\text{'s}}(b))\ldots)$$

can be defined by second-order primitive recursion:

$$\begin{aligned}
\rho_c(0, f(c), b) &= b \\
\rho_c(a+1, f(c), b) &= f(\rho_c(a, f(c), b))
\end{aligned}$$

Furthermore, we have two auxiliary functions

$$\iota(a,b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases} \quad \text{and} \quad \lambda(a,b) = \begin{cases} 0 & \text{if } a \neq b \\ 1 & \text{if } a = b \end{cases}$$

which are primitive recursive, as well as addition and multiplication. The term $\mathfrak{a}(a,b)$ is short for $\iota(a,1)\cdot\iota(a,0)\cdot b+\lambda(a,1)$; we then have

$$\mathfrak{a}(a,b)=\begin{cases}0 & \text{if } a=0\\ 1 & \text{if } a=1\\ b & \text{otherwise}\end{cases}$$

The Ackermann function is defined by

$$\begin{aligned}\varphi(0,b,c) &= b+c\\ \varphi(a+1,b,c) &= \rho_d(c,\varphi(a,b,d),\mathfrak{a}(a,b)).\end{aligned}$$

In more suggestive terms,

$$\begin{aligned}\varphi(0,b,c) &= b+c\\ \varphi(1,b,c) &= b\cdot c\\ \varphi(a+1,b,c) &= \underbrace{\varphi(a,b,\varphi(a,b,\ldots\varphi(a,b,b)\cdots))}_{c\text{ times}}\end{aligned}$$

The system of second-order primitive recursive arithmetic $2\mathrm{PRA}^-$ used in Ackermann (1924b) consists of axioms (1)–(15) of Hilbert and Bernays (1923b, see Section 2), axiom (16) was replaced by

16. $a\neq 0\rightarrow a=\delta(a)+1$,

plus defining equations for both first- and second-order primitive recursive functions. There is no induction rule (which is usually included in systems of primitive recursive arithmetic), although the consistency proof given by Ackermann can easily be extended to cover it.

3.2. The consistency proof for primitive recursive arithmetic

Allowing primitive recursion axioms for functions which contain function variables is a natural extension of the basic calculus of stages III and IIII, and is necessary in order to be able to introduce sufficiently complex functions. Hilbert seems to have thought that by extending primitive recursion in this way, or at least by building an infinite hierarchy of levels of primitive recursions using variables of higher types, he could account for *all* the number theoretic functions, and hence for all real numbers (represented as decimal expansions). In the spirit of the stage-by-stage development of systems of mathematics and consistency proofs, it is of course necessary to show the consistency of the system of stage IIII, which is the system presented by Ackermann. As before, it makes perfect sense to first establish the consistency for the fragment of stage IIII not containing the transfinite ε and π functions. In Section 4 of his dissertation, Ackermann undertakes precisely this aim.

The proof is a direct extension of the consistency proof of stage III, the elementary calculus of free variables with basic arithmetical axioms and primitive recursive definitions, i.e., PRA. This proof had already been presented in Hilbert's lectures in 1921–22 and 1922–23. The idea here is the same: put a given, purported proof of $0 \neq 0$ into tree form, eliminate variables, and reduce functionals. The remaining figure consists entirely of correct formulas, where correctness of a formula is a syntactically defined and easily decidable property. The only complication for the case where function variables are also admitted is the reduction of functionals. It must be shown that every functional, i.e., every term of the language, can be reduced to a numeral on the basis of the defining recursion equations. For the original case, this could be done by a relatively simple inductive proof. For the case of $2\mathrm{PRA}^-$, it is not so obvious.

Ackermann locates the difficulty in the following complication. Suppose you have a functional $\varphi_b(2, \mathfrak{b}(b))$. Here, $\mathfrak{b}(b)$ is a term which denotes a function, and so there is no way to replace the variable b with a numeral before evaluating the entire term. In effect, the variable b is bound (in modern notation, the term might be more suggestively written $\varphi(2, \lambda b.\mathfrak{b}(b))$.) In order to reduce this term, we apply the recursion equations for φ and might end up with a term like

$$\mathfrak{b}(1) + \mathfrak{b}(2) + \mathfrak{b}(0) \cdot \mathfrak{b}(1) + \mathfrak{b}(1) \cdot \mathfrak{b}(2).$$

The remaining $\mathfrak{b}$'s might in turn contain φ. By contrast, reducing a term $\psi(\mathfrak{z})$ where ψ is defined by first-order primitive recursion results in a term which does not contain ψ.

To show that the reduction indeed comes to an end if innermost subterms are reduced first, Ackermann proposes to assign indices to terms and show that each reduction reduces this index. The indices are, essentially, ordinal notations $< \omega^{\omega^{\omega}}$. Since this is probably the first proof using ordinal notations, it may be of some interest to repeat and analyze it in some detail here. In my presentation, I stay close to Ackermann's argument and only change the notation for ranks, indices, and orders: Where Ackermann uses sequences of natural numbers, I will use the more perspicuous ordinal notations.

Suppose the primitive recursive functions are arranged in a linear order according to the order of definition. We write $\varphi < \psi$ if φ occurs before ψ in the order of definition, i.e., ψ cannot be used in the defining equations for φ. Suppose further that we are given a specific term t. The notion of *subordination* is defined as follows: an occurrence of a function symbol ξ in t is subordinate to an occurrence of φ, if φ is the outermost function symbol of a subterm s, the occurrence of ξ is in s, and the subterm of s with outermost function symbol ξ contains a bound variable b in whose scope the occurrence of φ is (this includes the case where b happens to be bound by φ itself).[34] In

other words, $\mathfrak{t}$ is of the form

$$\mathfrak{t}'(\ldots\varphi_b(\ldots\xi(\ldots b\ldots)\ldots)\ldots)$$

The *rank* $rk(\mathfrak{t},\varphi)$ of an *occurrence* of a function symbol φ with respect to $\mathfrak{t}$ is defined as follows: If there is no occurrence of $\psi > \varphi$ which is subordinate to φ in $\mathfrak{t}$, then $rk(\mathfrak{t},\varphi) = 1$. Otherwise,

$$rk(\mathfrak{t},\varphi) = \max\{rk(\mathfrak{t},\psi) : \psi > \varphi \text{ is subordinate to } \varphi\} + 1.$$

The rank $r(\mathfrak{t},\varphi)$ of a term $\mathfrak{t}$ with respect to a function symbol φ is the maximum of the ranks of occurrences of φ or $\psi > \varphi$ in t. (If neither φ nor $\psi > \varphi$ occur in $\mathfrak{t}$, that means $r(\mathfrak{t},\varphi) = 0$.

We assign to a subterm $\mathfrak{s}$ of $\mathfrak{t}$ a sequence of ranks of ψ_n, …, ψ_0 with respect to $\mathfrak{s}$, where $\psi_0 < \cdots < \psi_n$ are all function symbols occurring in $\mathfrak{t}$. This is the *order* of $\mathfrak{s}$:

$$o(\mathfrak{s}) = \langle r(\mathfrak{s},\psi_n),\ldots,r(\mathfrak{s},\psi_0)\rangle$$

We may think of this order as an ordinal $< \omega^\omega$, specifically, $o(\mathfrak{s})$ corresponds to

$$\omega^n \cdot r(\mathfrak{s},\psi_n)\cdot + \cdots + \omega\cdot r(\mathfrak{s},\psi_1)\cdot + r(\mathfrak{s},\psi_0)$$

Now consider the set of all distinct subterms of t of a given order o which are not numerals. The *degree* $d(\mathfrak{t},o)$ of o in $\mathfrak{t}$ is the cardinality of that set. The *index* $j(\mathfrak{t})$ of $\mathfrak{t}$ is the sequence of degrees ordered in the same way as the orders, i.e., it corresponds to an ordinal of the form

$$\sum_o \omega^o \cdot d(\mathfrak{t},o)$$

where the sum ranges over all orders o. Obviously, this is an ordinal $< \omega^{\omega^\omega}$.

Ackermann does not use ordinals to define these indices; he stresses that he is only dealing with finite sequences of numbers, on which an elaborate order is imposed. Rather than appeal to the well-orderedness of ω^{ω^ω}, he gives a more direct argument that by repeatedly proceeding to indices which are smaller in the imposed order one eventually has to reach the index which consists of all 0.

Suppose a term $\mathfrak{t}$ not containing ε or π is given. Let $\mathfrak{s}$ be an innermost constant subterm which is not a numeral, we may assume it is of the form $\varphi_{\bar{b}}(\mathfrak{z}_1,\ldots,\mathfrak{z}_n,\mathfrak{u}_1,\ldots,\mathfrak{u}_m)$ where $\mathfrak{u}_i$ is a term with at least one variable bound by φ and which doesn't contain a constant subterm. We have two cases:

(1) $\mathfrak{s}$ does not contain bound variables, i.e., $m = 0$. The order of $\mathfrak{s}$ is ω^k (where $\varphi = \psi_k$). Evaluating the term $\mathfrak{s}$ by recursion results in a term $\mathfrak{s}'$ in which only function symbols of lower index occur. Hence, the highest exponent in the order of $\mathfrak{s}'$ is less than i, and so $o(s') < o(s)$. Furthermore, since

no variable which is bound in t can occur in $\mathfrak{s}'$ (since no such variable occurs in $\mathfrak{s}$), replacing $\mathfrak{s}$ by $\mathfrak{s}'$ in $\mathfrak{t}$ does not result in new occurrences of function symbols which are subordinate to any other. Thus the number of subterms in the term $\mathfrak{t}'$ which results from such a replacement with orders $> o(\mathfrak{s})$ remains the same, while the number of subterms of order $o(\mathfrak{s})$ is reduced by 1. Hence, $j(\mathfrak{t}') < j(\mathfrak{t})$.

(2) $\mathfrak{s}$ does contain bound variables. For simplicity, assume that there is one numeral argument and one functional argument, i.e., $\mathfrak{s}$ is of the form $\varphi_b(\mathfrak{z}, \mathfrak{c}(b))$. In this case, all function symbols occurring in $\mathfrak{c}(b)$ are subordinate to φ, or otherwise $\mathfrak{c}(b)$ would contain a constant subterm.[35] Thus, the rank of $\mathfrak{c}(b)$ in $\mathfrak{t}$ with respect to ψ_i is less that the rank of $\mathfrak{s}$ with respect to ψ_i.

We reduce the subterm $\mathfrak{s}$ to a subterm $\mathfrak{s}'$ by applying the recursion. $\mathfrak{s}'$ does not contain the function symbol φ. We want to show that replacing $\mathfrak{s}$ by $\mathfrak{s}'$ in $\mathfrak{t}$ lowers the index of $\mathfrak{t}$.

First, note that when substituting a term $\mathfrak{a}$ for b in $\mathfrak{c}(b)$, the order of the resulting $\mathfrak{c}(\mathfrak{a})$ with respect to φ is the maximum of the orders of $\mathfrak{c}(b)$ and $\mathfrak{a}$, since none of the occurrences of function symbols in $\mathfrak{a}$ contain bound variables whose scope begins outside of $\mathfrak{a}$, and so none of these variables are subordinate to any function symbols in $\mathfrak{c}(b)$.

Now we prove the claim by induction on $\mathfrak{z}$. Suppose the defining equation for φ is

$$\begin{aligned} \varphi_b(0, f(b)) &= \mathfrak{a}_b(f(b)) \\ \varphi_b(a+1, f(b)) &= \mathfrak{b}_b(\varphi_c(a, f(c)), a, f(b)). \end{aligned}$$

If $\mathfrak{z} = 0$, then $\mathfrak{s}' = \mathfrak{a}_b(\mathfrak{c}(b))$. At a place where $f(b)$ is an argument to a function, $f(b)$ is replaced by $\mathfrak{c}(d)$, and d is not in the scope of any φ (since $\mathfrak{a}$ doesn't contain φ). For instance, $\mathfrak{a}_b(f(b)) = 2 + \psi_d(3, f(d))$. Such a replacement cannot raise the φ-rank of $\mathfrak{s}'$ above that of $\mathfrak{c}(b)$. The term $\mathfrak{c}$ might also be used in places where it is not a functional argument, e.g., if $\mathfrak{a}_b(f(b)) = f(\psi(f(2)))$. By a simple induction on the nesting of f's in $\mathfrak{a}_b(f(b))$ it can be seen that the φ-rank of $\mathfrak{s}'$ is the same as that of $\mathfrak{c}(b)$: For $\mathfrak{c}(\mathfrak{d})$ where $\mathfrak{d}$ does not contain $\mathfrak{c}$, the φ-rank of $\mathfrak{c}(\mathfrak{d})$ equals that of $\mathfrak{c}(b)$ by the note above and the fact that $\mathfrak{d}$ does not contain φ. If $\mathfrak{d}$ does contain a nested $\mathfrak{c}$, then by induction hypothesis and the first case, its φ-rank is the same as that of $\mathfrak{c}(b)$. By the note, again, the entire subterm has the same φ-rank as $\mathfrak{c}(b)$.

The case of $\varphi_b(\mathfrak{z}+1, \mathfrak{c}(b))$ is similar. Here, the first replacement is

$$\mathfrak{b}_b(\varphi_c(\mathfrak{z}, \mathfrak{c}(c)), \mathfrak{z}, \mathfrak{c}(b)).$$

Further recursion replaces $\varphi_c(\mathfrak{z}, \mathfrak{c}(c))$ by another term which, by induction hypothesis, has φ-rank less than or equal to that of $\mathfrak{c}(b)$. The same considerations as in the base case show that the entire term also has a φ-rank no larger than $\mathfrak{c}(b)$.

We have thus shown that eliminating the function symbol φ by recursion from an innermost constant term reduces the φ-rank of the term at least by one and does not increase the ψ_j-ranks of any subterms for any $j > i$.

In terms of ordinals, this shows that at least one subterm of order o was reduced to a subterm of order $o' < o$, all newly introduced subterms have order $< o$, and the order of no old subterm increased. Thus, the index of the entire term was reduced. The factor $\omega^o \cdot n$ changed to $\omega^o \cdot (n-1)$.

> We started with a given constant function, which we characterized by a determinate index. We replaced a $\varphi_b(\mathfrak{z}, \mathfrak{c}(b))$ within that functional by another functional, where the φ-rank decreased and the rank with respect to functions to the right of φ [i.e., which come after φ in the order of definition] did not increase. Now we apply the same operation to the resulting functional. After finitely many steps we obtain a functional which contains no function symbols at all, i.e., it is a numeral.
>
> We have thus shown: a constant functional, which does not contain ε and π, can be reduced to a numeral in finitely many steps.[36]

3.3. Ordinals, transfinite recursion, and finitism

It is quite remarkable that the earliest extensive and detailed technical contribution to the finitist project would make use of transfinite induction in a way not dissimilar to Gentzen's later proof by induction up to ε_0. This bears on a number of questions regarding Hilbert's understanding of the strength of finitism. In particular, it is often said that Gentzen's proof is not finitist, because it uses transfinite induction. However, Ackermann's original consistency proof for $2\mathrm{PRA}^-$ also uses transfinite induction, using an index system which is essentially an ordinal notation system, just like Gentzen's. If it is granted that Ackermann's proof is finitistic, but Gentzen's is not, i.e., transfinite induction up to ω^{ω^ω} is finitistic but not up to ε_0, then where—and why—should the line be drawn? Furthermore, the consistency proof of $2\mathrm{PRA}^-$ is in essence a—putatively finitistic—explanation of how to compute second order primitive recursive functions, and a proof that the computation procedure defined by them always terminates. In other words, it is a finitistic proof that second order primitive recursive functions are well defined.[37]

Ackermann was completely aware of the involvement of transfinite induction in this case, but he sees in it no violation of the finitist standpoint.

> The disassembling of functionals by reduction does not occur in the sense that a finite ordinal is decreased each time an outermost function symbol is eliminated. Rather, to each functional corresponds as it were a transfinite ordinal number as its rank, and the theorem, that a constant functional is reduced to a numeral after carrying out finitely many operations, corresponds to the other [theorem], that if one descends from a transfinite ordinal number to ever smaller ordinal numbers, one has to

> reach zero after a finite number of steps. Now there is naturally no mention of transfinite sets or ordinal numbers in our metamathematical investigations. It is however interesting, that the mentioned theorem about transfinite ordinals can be formulated so that there is nothing transfinite about it.
>
> Consider a transfinite ordinal number less than $\omega \cdot \omega$. Each such ordinal number can be written in the form: $\omega \cdot \mathfrak{n} + \mathfrak{m}$, where $\mathfrak{n}$ and $\mathfrak{m}$ are finite numbers. Hence such an ordinal can also be characterized by a pair of finite numbers $(\mathfrak{n}, \mathfrak{m})$, where the order of these numbers is of course significant. To the descent in the series of ordinals corresponds the following operation on the number pair $(\mathfrak{n}, \mathfrak{m})$. Either the first number $\mathfrak{n}$ remains the same, then the number $\mathfrak{m}$ is replaced by a smaller number $\mathfrak{m}'$. Or the first number $\mathfrak{n}$ is made smaller; then I can put an arbitrary number in the second position, which can also be larger than $\mathfrak{m}$. It is clear that one has to reach the number pair $(0,0)$ after finitely many steps. For after at most $\mathfrak{m}+1$ steps I reach a number pair, where the first number is smaller than $\mathfrak{n}$. Let $(\mathfrak{n}', \mathfrak{m}')$ be that pair. After at most $\mathfrak{m}'+1$ steps I reach a number pair in which the first number is again smaller than $\mathfrak{n}'$, etc. After finitely many steps one reaches the number pair $(0,0)$ in this fashion, which corresponds to the ordinal number 0. In this form, the mentioned theorem contains nothing transfinite whatsoever; only considerations which are acceptable in metamathematics are used. The same holds true if one does not use pairs but triples, quadruples, etc. This idea is not only used in the following proof that the reduction of functionals terminates, but will also be used again and again later on, especially in the finiteness proof at the end of the work.[38]

Over ten years later, Ackermann discusses the application of transfinite induction for consistency proofs in correspondence with Bernays. Gentzen's consistency proof had been published (Gentzen, 1936), and Gentzen asks, through Bernays,

> whether you [Ackermann] think that the method of proving termination [*Endlichkeitsbeweis*] by transfinite induction can be applied to the consistency proof of your dissertation. I would like it very much, if that were possible.[39]

In his reply, Ackermann recalls his own use of transfinite ordinals in the 1924 dissertation.

> I just realized now, as I am looking at my dissertation, that I operate with transfinite ordinals in a similar fashion as Gentzen.[40]

A year and a half later, Ackermann mentions the transfinite induction used in his dissertation again:

> I do not know, by the way, whether you are aware (I did at the time not consider it as a transgression beyond the narrower finite standpoint), that

> transfinite inferences are used in my dissertation. (Cf., e.g., the remarks in the last paragraph on page 13 and the following paragraph of my dissertation).[41]

These remarks may be puzzling, since they seem to suggest that Bernays was not familiar with Ackermann's work. This is clearly not the case. Bernays corresponded with Ackerman extensively in the mid-20s about the ε-substitution method and the decision problem, and had clearly studied Ackermann's dissertation. Neither Bernays nor Hilbert are on record objecting to the methods used in Ackermann's dissertation. It can thus be concluded that Ackermann's use of transfinite induction was considered acceptable from the finitist standpoint.

3.4. THE ε-SUBSTITUTION METHOD

As we have seen above, Hilbert had outlined an idea for a consistency proof for systems involving ε-terms already in early 1922 (Hilbert, 1922a), and a little more precisely in his talk of 1922 (Hilbert, 1923) and in the 1922–23 lectures (Hilbert and Bernays, 1923a). Let us review the *Ansatz* in the notation used in 1924: Suppose a proof involves only one ε term $\varepsilon_a A(a)$ and corresponding *critical formulas*

$$A(\mathfrak{k}_i) \to A(\varepsilon_a A(a)),$$

i.e., substitution instances of the transfinite axiom

$$A(a) \to A(\varepsilon_a A(a)).$$

We replace $\varepsilon_a A(a)$ everywhere with 0, and transform the proof as before by rewriting it in tree form ("dissolution into proof threads"), eliminating free variables and evaluating numerical terms involving primitive recursive functions. Then the critical formulas take the form

$$A(\mathfrak{z}_i) \to A(0),$$

where $\mathfrak{z}_i$ is the numerical term to which $\mathfrak{k}_i$ reduces. A critical formula can now only be false if $A(\mathfrak{z}_i)$ is true and $A(0)$ is false. If that is the case, repeat the procedure, now substituting $\mathfrak{z}_i$ for $\varepsilon_a A(a)$. This yields a proof in which all initial formulas are correct and no ε terms occur.

If critical formulas of the second kind, i.e., substitution instances of the induction axiom,

$$\varepsilon_a A(a) \neq 0 \to \overline{A(\delta\varepsilon_a A(a))},$$

also appear in the proof, the witness $\mathfrak{z}$ has to be replaced with the least $\mathfrak{z}'$ so that $A(\mathfrak{z}')$ is true.

The challenge was to extend this procedure to (a) cover more than one ε-term in the proof, (b) take care of nested ε-terms, and lastly (c) extend it

to second-order ε's and terms involving them, i.e, $\varepsilon_f \mathfrak{A}_a(f(a))$. This is what Ackermann set out to do in the last part of his dissertation, and what he and Hilbert thought he had accomplished.[42]

The system for which Ackermann attempted to give a consistency proof consisted of the system of second-order primitive recursive arithemtic (see Section 3.1 above) together with the transfinite axioms:

1. $A(a) \to A(\varepsilon_a A(a))$ $\qquad A_a f(a) \to A_a((\varepsilon_f A_b(f(b)))(a))$
2. $A(\varepsilon_a A(a)) \to \pi_a A(a) = 0$ $\qquad A_a(\varepsilon_f A_b(f(b))(a)) \to \pi_f A_a(f(a)) = 0$
3. $\overline{A(\varepsilon_a A(a))} \to \pi_a A(a) = 1$ $\qquad \overline{A_a(\varepsilon_f A_b(f(b))(a))} \to \pi_f A_a(f(a)) = 1$
4. $\varepsilon_a A(a) \neq 0 \to \overline{A(\delta(\varepsilon_a A(a)))}$[43]

The intuitive interpretation of ε and π, based on these axioms, is obvious: $\varepsilon_a \mathfrak{A}(a)$ is a witness for $\mathfrak{A}(a)$ if one exists, and $\pi_a \mathfrak{A}(a) = 1$ if $\mathfrak{A}(a)$ is false for all a, and $= 0$ otherwise. The π functions are not necessary for the development of mathematics in the axiom system. They do, however, serve a function in the consistency proof, viz., to keep track of whether a value of 0 for $\varepsilon_a \mathfrak{A}(a)$ is a "default value" (i.e., a trial substitution for which $\mathfrak{A}(a)$ may or may not be true) or an actual witness (a value for which $\mathfrak{A}(a)$ has been found to be true).

I shall now attempt to give an outline of the ε-substitution procedure defined by Ackermann. For simplicity, I will leave the case of second-order ε-terms (i.e., those involving ε_f) to the side.

An *ε-term* is an expression of the form $\varepsilon_a \mathfrak{A}(a)$, where a is the only free variable in $\mathfrak{A}$, and similarly for a π-term. For the purposes of the discussion below, we will not specifically refer to π's unless necessary, and most definitions and operations apply equally to ε-terms and π-terms. If a formula $A(a)$ or an ε-term $\varepsilon_a \mathfrak{A}(a)$ contains no variable-free subterms which are not numerals, we call them *canonical*. Canonical formulas and ε-terms are indicated by a tilde: $\varepsilon_a \tilde{\mathfrak{A}}(a)$.

The main notion in Ackermann's proof is that of a *total substitution* S (*Gesamtersetzung*). It is a mapping of canonical ε- and π-terms to numerals and 0 or 1, respectively. When canonical ε-terms in a proof are successively replaced by their values under the mapping, a total substitution reduces the proof to one not containing any ε's. If S maps $\varepsilon_a \tilde{\mathfrak{A}}(a)$ to $\mathfrak{z}$ and $\pi_a \tilde{\mathfrak{A}}(a)$ to i, then we say that $\tilde{\mathfrak{A}}(a)$ receives a $(\mathfrak{z}, \mathfrak{i})$ substitution under S and write $S(\tilde{\mathfrak{A}}(a)) = (\mathfrak{z}, \mathfrak{i})$.

It is of course not enough to define a mapping from the canonical ε-terms *occuring in the proof* to numerals: The proof may contain, e.g., $\varepsilon_a \mathfrak{A}(a, \varphi(\varepsilon_b \mathfrak{B}(b)))$. To reduce this to a numeral, we first need a value $\mathfrak{z}$ for the term $\varepsilon_b \mathfrak{B}(b)$. Replacing $\varepsilon_b \mathfrak{B}(b)$ by $\mathfrak{z}$, we obtain $\varepsilon_a \mathfrak{A}(a, \varphi(\mathfrak{z}))$. Suppose the value $\varphi(\mathfrak{z})$ is $\mathfrak{z}'$. The total substitution then also has to specify a substitution for $\varepsilon_a \mathfrak{A}(a, \mathfrak{z}')$.

Given a total substitution S, a proof is reduced to an ε-free proof as follows: First all ε-free terms are evaluated. (Such terms contain only numerals and primitive recursive functions; these are computed and the term replaced by the numeral corresponding to the value of the term) Now let $\mathfrak{e}_{11}$, $\mathfrak{e}_{12}$, ... be all the innermost (canonical) ε- or π-terms in the proof, i.e., all ε- or π-terms which do not themselves contain nested ε- or π-terms or constant (variable-fee) subterms which are not numerals. The total substitution specifies a numeral substitution for each of these. Replace each $\mathfrak{e}_{1i}$ by its corresponding numeral. Repeat the procedure until the only remaining terms are numerals. We write $|\mathfrak{e}|_S$ for the result of applying this procedure to the expression (formula or term) $\mathfrak{e}$. Note that $|\mathfrak{e}|_S$ is canonical.

Based on this reduction procedure, Ackermann defines a notion of subordination of canonical formulas. Roughly, a formula $\tilde{\mathfrak{B}}(b)$ is subordinate to $\tilde{\mathfrak{A}}(a)$ if in the process of reducing some formula $\tilde{\mathfrak{A}}(\mathfrak{z})$, an ε-term $\varepsilon_b\tilde{\mathfrak{B}}(b)$ is replaced by a numeral. For instance, $a = b$ is subordinate to $a = \varepsilon_b(a = b)$. Indeed, if $\tilde{\mathfrak{A}}(a)$ is $a = \varepsilon_b(a = b)$, then the reduction of $\tilde{\mathfrak{A}}(\mathfrak{z}) \equiv \mathfrak{z} = \varepsilon_b(\mathfrak{z} = b)$ would use a replacement for the ε-term belonging to $\tilde{\mathfrak{B}}(\mathfrak{z} = b)$.[44] It is easy to see that this definition corresponds to the notion of subordination as defined in Hilbert and Bernays (1939). An ε-*expression* is an expression of the form $\varepsilon_a\mathfrak{A}(a)$. If $\varepsilon_a\mathfrak{A}(a)$ contains no free variables, it is called an ε-*term*. If an ε-term $\varepsilon_b\mathfrak{B}(b)$ occurs in an expression (and is different from it), it is said to be *nested* in it. If an ε-expression $\varepsilon_b\mathfrak{B}(a,b)$ occurs in an expression in the scope of ε_a, then it is *subordinate* to that expression. Accordingly, we can define the degree of an ε-term and the rank of an ε-expression as follows: An ε-term with no nested ε-subterms is of degree 1; otherwise its degree is the maximum of the degrees of its nested ε-subterms $+1$. The rank of an ε-expression with no subordinate ε-expressions is 1; otherwise it is the maximum of the ranks of its subordinate ε-expressions $+1$. If $\tilde{\mathfrak{B}}(b)$ is subordinate to $\tilde{\mathfrak{A}}(a)$ according to Ackermann's definition, then $\varepsilon_b\tilde{\mathfrak{B}}(b)$ is subordinate in the usual sense to $\varepsilon_a\tilde{\mathfrak{A}}(a)$, and the rank of $\varepsilon_b\tilde{\mathfrak{B}}(b)$ is less than that of $\varepsilon_a\tilde{\mathfrak{A}}(a)$. The notion of degree corresponds to an ordering of canonical formulas used for the reduction according to a total substitution in Ackermann's procedural definition: First all ε-terms of degree 1 (i.e., all innermost ε-terms) are replaced, resulting (after evaluation of primitive recursive functions) in a partially reduced proof. The formulas corresponding to innermost ε-terms now are reducts of ε-terms of degree 2 in the original proof. The canonical formulas corresponding to ε-terms of degree 1 are called the formulas of *level* 1, the canonical formulas corresponing to the innermost ε-terms in the results of the first reduction step are the formulas of level 2, and so forth.

The consistency proof proceeds by constructing a sequence S_1, S_2, ... of total substitutions together with bookkeeping functions $f_i(\tilde{\mathfrak{A}}(a), j) \to \{0,1\}$,[45] which eventually results in a *solving substitution*, i.e., a total substitution which reduces the proof to one which contains only correct ε-free formulas.

We begin with a total substitution S_1 which assigns $(0,1)$ to all canonical formulas, and set $f_1(\tilde{\mathfrak{A}}(a),1) = 1$ for all $\tilde{\mathfrak{A}}(a)$ for which S_1 assigns a value. If S_i is a solving substitution, the procedure terminates. Otherwise, the next total substitution S_{i+1} is obtained as follows: If S_i is not a solving substitution, at least one of the critical formulas in the proof reduces to an incorrect formula. We have three cases:

1. Either an ε-axiom $\mathfrak{A}(a) \to \mathfrak{A}(\varepsilon_a \mathfrak{A}(a))$ or a π-axiom of the first kind $\mathfrak{A}(\varepsilon_a \mathfrak{A}(a)) \to \pi_a \mathfrak{A}(a) = 0$ reduces to a false formula of the form $\tilde{\mathfrak{A}}(\mathfrak{z}) \to \tilde{\mathfrak{A}}(0)$ or $\tilde{\mathfrak{A}}(0) \to 1 = 0$, and $S_i(\tilde{\mathfrak{A}}(a)) = (0,1)$. Pick one such $\tilde{\mathfrak{A}}(a)$ of lowest level (i.e., $\varepsilon_a \mathfrak{A}(a)$ of lowest degree).

   If $\tilde{\mathfrak{A}}(0) \to 1 = 0$ is incorrect, $\tilde{\mathfrak{A}}(0)$ must be correct; let $S_{i+1}(\tilde{A}(a)) = (0,0)$. Otherwise $\tilde{\mathfrak{A}}(\mathfrak{z}) \to \tilde{\mathfrak{A}}(0)$ is incorrect and hence $\tilde{\mathfrak{A}}(\mathfrak{z})$ must be correct; then let $S_{i+1}(\tilde{\mathfrak{A}}(a)) = (\mathfrak{z},0)$. In either case, set $f_{i+1}(\tilde{\mathfrak{A}}(a), i+1) = 1$.

   For other formulas $\tilde{\mathfrak{B}}(b)$, $S_{i+1}(\tilde{\mathfrak{B}}(b)) = S_j(\tilde{\mathfrak{B}}(b))$ where j is the greatest index $\leq i$ such that $f_i(\tilde{\mathfrak{B}}(b), j) = 1$. $S_{i+1}(\tilde{\mathfrak{B}}(b)) = (0,1)$ if no such j exists (i.e., $\tilde{\mathfrak{B}}(b)$ has never before received an example substitution). Also, let $f_{i+1}(\tilde{\mathfrak{B}}(b), i+1) = 1$. For all canonical formulas $\tilde{\mathfrak{C}}(c)$, let $f_{i+1}(\tilde{\mathfrak{C}}(c), j) = f_i(\tilde{\mathfrak{C}}(c), j)$ for $j \leq i$.

2. Case (1) does not apply, but at least one of the minimality axioms $\varepsilon_a \mathfrak{A}(a) \neq 0 \to \overline{\mathfrak{A}}(\delta(\varepsilon_a \mathfrak{A}(a))$ reduces to a false formula, $\mathfrak{z} \neq 0 \to \overline{\tilde{\mathfrak{A}}}(\mathfrak{z}-1)$. This is only possible if $S_i(\tilde{\mathfrak{A}}(a)) = (\mathfrak{z},0)$. Again, pick the one of lowest level, and let $S_{i+1}(\tilde{\mathfrak{A}}(a)) = (\mathfrak{z}-1,0)$ and $f_{i+1}(\tilde{\mathfrak{A}}(a), i+1) = 1$. Substitutions for other formulas and bookkeeping functions are defined as in case (1).

3. Neither case (1) nor (2) applies, but some instance of an ε-axiom of the form $\mathfrak{A}(a) \to \mathfrak{A}(\varepsilon_a \mathfrak{A}(a))$ or of a π-axiom of the form $\overline{\mathfrak{A}}(\varepsilon_a \mathfrak{A}(a)) \to \pi_a \mathfrak{A}(a) = 1$, e.g., $\tilde{\mathfrak{A}}(\mathfrak{a}) \to \tilde{\mathfrak{A}}(\mathfrak{z}))$ or $\overline{\tilde{\mathfrak{A}}}(\mathfrak{z}) \to 0 = 1$, reduces to an incorrect formula. We then have $S_i(\tilde{\mathfrak{A}}(a)) = (\mathfrak{z},0)$ (since otherwise case (1) would apply). In either case, $|\tilde{\mathfrak{A}}(\mathfrak{z}))|_{S_i}$ must be incorrect. Let j be the least index where $S_j(\tilde{\mathfrak{A}}(a)) = (\mathfrak{z},0)$ and $f_i(\tilde{\mathfrak{A}}(a)), j) = 1$. At the preceding total substitution S_{j-1}, $S_j(\tilde{\mathfrak{A}}(a)) = (0,1)$ or $S_{j-1}(A(a)) = (\mathfrak{z}+1,0)$, and $|\mathfrak{A}(\mathfrak{z})|_{S_{j-1}}$ is correct. $\tilde{\mathfrak{A}}(\mathfrak{z})$ thus must reduce to different formulas under S_{j-1} and under S_i, which is only possible if a formula subordinate to $\tilde{\mathfrak{A}}$ reduces differently under S_{j-1} and S_i.

   For example, suppose $\tilde{\mathfrak{A}}(a)$ is really $\tilde{\mathfrak{A}}(a, \varepsilon_b \tilde{\mathfrak{B}}(a,b))$. Then the corresponding ε-axiom would be

$$\tilde{\mathfrak{A}}(a, \varepsilon_b \tilde{\mathfrak{B}}(a,b)) \to \tilde{\mathfrak{A}}(\underbrace{\varepsilon_a \tilde{\mathfrak{A}}(a, \varepsilon_b \tilde{\mathfrak{B}}(a,b))}_{\varepsilon_a \tilde{\mathfrak{A}}(a)}, \varepsilon_b \tilde{\mathfrak{B}}(\underbrace{\varepsilon_a \tilde{\mathfrak{A}}(a, \varepsilon_b \tilde{\mathfrak{B}}(a,b))}_{\varepsilon_a \tilde{\mathfrak{A}}(a)}, b))$$

An instance thereof would be

$$\tilde{\mathfrak{A}}(\mathfrak{a}, \varepsilon_b \tilde{\mathfrak{B}}(\mathfrak{a}, b) \to \tilde{\mathfrak{A}}(\varepsilon_a \tilde{\mathfrak{A}}(a, \varepsilon_b \tilde{\mathfrak{B}}(a,b)), \varepsilon_b \tilde{\mathfrak{B}}(\varepsilon_a \tilde{\mathfrak{A}}(a, \varepsilon_b \tilde{\mathfrak{B}}(a,b)), b)).$$

This formula, under a total substitution with $S_i(\tilde{\mathfrak{A}}(a, \varepsilon_b \tilde{\mathfrak{B}}(a,b))) = (\mathfrak{z}, 0)$ reduces to

$$\tilde{\mathfrak{A}}(\mathfrak{a}, \varepsilon_b \tilde{\mathfrak{B}}(\mathfrak{a}, b) \to \tilde{\mathfrak{A}}(\mathfrak{z}, \varepsilon_b \tilde{\mathfrak{B}}(\mathfrak{z}, b))$$

The consequent of this conditional, i.e., $\tilde{\mathfrak{A}}(\mathfrak{z})$, can reduce to different formulas under S_i and S_{j-1} only if $\varepsilon_b \tilde{\mathfrak{B}}(\mathfrak{z}, b)$ receives different substitutions under S_i and S_{j-1}, and $\tilde{\mathfrak{B}}(a,b)$ is subordinate to $\tilde{\mathfrak{A}}(a)$.

The next substitution is now defined as follows: Pick an innermost formula subordinate to $\tilde{\mathfrak{A}}(a)$ which changes substitutions, say $\tilde{\mathfrak{B}}(b)$. For all formulas $\tilde{\mathfrak{C}}(c)$ which are subordinate to $\tilde{\mathfrak{B}}(b)$ as well as $\tilde{\mathfrak{B}}(b)$ itself, we set $f_{i+1}(\tilde{\mathfrak{C}}(c), k) = 1$ for $j \le k \le i+1$ and $f_{i+1}(\tilde{\mathfrak{C}}(c), k) = 0$ for all other formulas. For $k < j$ we set $f_{i+1}(\tilde{\mathfrak{C}}(c), n) = f_i(\tilde{\mathfrak{C}}(c), k)$ for all $\tilde{\mathfrak{C}}(c)$. The next substitution S_{i+1} is now given by $S_{i+1}(\tilde{\mathfrak{C}}(c)) = S_k(\tilde{\mathfrak{C}}(c))$ for k greatest such that $f_{i+1}(\tilde{\mathfrak{C}}(c), k) = 1$ or $= (0,1)$ if no such k exists.

Readers familiar with the substitution method defined in Ackermann (1940) will note the following differences:

a. Ackermann (1940) uses the notion of a *type* of an ε-term and instead of defining total substitutions in terms of numeral substitutions for canonical ε-terms, assigns a function of finite support to ε-types. This change is merely a notational convenience, as these functional substitutions can be recovered from the numeral substitutions for canonical ε-terms. For example, if S assigns the substitutions to the canonical terms on the left, then a total substitution in the sense of Ackermann (1940) would assign the function g on the right to the type $\varepsilon_a \tilde{\mathfrak{A}}(a,b)$:

$$\begin{array}{ll} S(\tilde{\mathfrak{A}}(a, \mathfrak{z}_1) = \mathfrak{z}'_1 & \quad g(\varepsilon_a \tilde{\mathfrak{A}}(a,b))(\mathfrak{z}_1) = \mathfrak{z}'_1 \\ S(\tilde{\mathfrak{A}}(a, \mathfrak{z}_3) = \mathfrak{z}'_2 & \quad g(\varepsilon_a \tilde{\mathfrak{A}}(a,b))(\mathfrak{z}_2) = \mathfrak{z}'_2 \\ S(\tilde{\mathfrak{A}}(a, \mathfrak{z}_2) = \mathfrak{z}'_3 & \quad g(\varepsilon_a \tilde{\mathfrak{A}}(a,b))(\mathfrak{z}_3) = \mathfrak{z}'_3 \end{array}$$

b. In case (2), dealing with the least number (induction) axiom, the next substitution is defined by reducing the substituted numeral $\mathfrak{z}$ by 1, whereas in (Ackermann, 1940), we immediately proceed to the least $\mathfrak{z}'$ such that $\tilde{\mathfrak{A}}(\mathfrak{z}')$ is correct. This makes the procedure converge more slowly, but also suggests that in certain cases (depending on which other critical formulas occur in the proof), the solving substitution does not necessarily provide example substitutions which are, in fact, least witnesses.

c. The main difference in the method lies in case (3). Whereas in (1940), example substitutions for all ε-types of rank lower than that of the changed

$\varepsilon_a \tilde{\mathfrak{A}}(a)$ are retained, and all others are reset to initial substitutions (functions constant equal to 0), in (1924b), only the substitutions of some ε-terms actually subordinate to $\varepsilon_a \tilde{\mathfrak{A}}(a)$ are retained, while others are not reset to initial substitutions, but to substitutions defined at some previous stage.

3.5. ASSESSMENT AND COMPLICATIONS

A detailed analysis of the method and of the termination proof given in the last part of Ackermann's dissertation must wait for another occasion, if only for lack of space. A preliminary assessment can, however, already be made on the basis of the outline of the substitution process above. Modulo some needed clarification in the definitions, the process is well-defined and terminates at least for proofs containing only least-number axioms (critical formulas corresponding to axiom (4)) of rank 1. The proof that the procedure terminates (§9 of Ackermann (1924b)) is opaque, especially in comparison to the proof by transfinite induction for primitive recursive arithmetic. The definition of a substitution method for second-order ε-terms is insufficient, and in hindsight it is clear that a correct termination proof for this part could not have been given with the methods available.[46]

Leaving aside, for the time being, the issue of what was *actually* proved in Ackermann (1924b), the question remains of what was *believed* to have been proved at the time. The system, as given in (1924b), had two major shortcomings: A footnote, added in proof, states:

> [The formation of ε-terms] is restricted in that a function variable $f(a)$ may not be substituted by a functional $\mathfrak{a}(a)$, in which a occurs in the scope of an ε_f.[47]

This applies in particular to the second-order ε-axioms

$$A_a(f(a)) \to A_a(\varepsilon_f A_b(f(b))(a)).$$

If we view $\varepsilon_f A_b(f(b))(a)$ as the function "defined by" A, and hence the ε-axiom as the ε-calculus analog of the comprehension axiom, this amounts roughly to a restriction to arithmetic comprehension, and thus a predicative system. This shortcoming, and the fact that the restriction turns the system into a system of predicative mathematics was pointed out by von Neumann (1927).

A second lacunae was the omission of an axiom of ε-extensionality for second-order ε-terms, i.e.,

$$(\forall f)(A(f) \rightleftharpoons B(f)) \to \varepsilon_f A(f) = \varepsilon_f B(f),$$

which corresponds to the axiom of choice. Both problems were the subject of correspondence with Bernays in 1925.[48] A year later, Ackermann is still trying to extend and correct the proof, now using ε-types:

> I am currently working again on the ε_f-proof and am pushing hard to finish it. I have already told you that the problem can be reduced to one of number theory. To prove the number-theoretic theorem seems to me, however, equally hard as the problem itself. I am now again taking the approach, which I have tried several times previously, to extend the definition of a ground type so that even ε with free function variables receive a substitution. This approach seems to me the most natural, and the equality axioms $(f)(A(f) \rightleftharpoons B(f)) \to \varepsilon_f A f \equiv \varepsilon_f B f$ would be treated simultaneously. I am hopeful that the obstacles previously encountered with this method can be avoided, if I use the ε_a formalism and use substitutions for the ε_f which may contain ε_a instead of functions defined without ε. I have, however, only thought through some simple special cases.[49]

In 1927, Ackermann developed a second proof of ε-substitution, using some of von Neumann's ideas (in particular, the notion of an ε-type, *Grundtyp*). The proof is unfortunately not preserved in its entirety, but references to it can be found in the correspondence. On April 12, 1927, Bernays writes to Ackermann:

> Finally I have thought through your newer proof for consistency of the ε_a's based on what you have written down for me befor your departure, and believe that I have seen the proof to be correct.[50]

Ackermann also refers to the proof in a letter to Hilbert from 1933:

> As you may recall, I had at the time a second proof for the consistency of the ε_a's. I never published that proof, but communicated it to Prof. Bernays orally, who then verified it. Prof. Bernays wrote to me last year that the result does not seem to harmonize with the work of Gödel.[51]

In Hilbert's address to the International Congress of Mathematicians in 1928 (Hilbert, 1928), the success of Ackermann's and von Neumann's work on ε-substitution for first-order systems is also taken for granted. Although Hilbert poses the extension of the proof to second-order systems as an open problem, there seems no doubt in his mind that the solution is just around the corner.[52]

It might be worthwhile to mention at this point that at roughly the same time a third attempt to find a satisfactory consistency proof was made. This attempt was based not on ε-substituion, but on Hilbert's so-called unsuccessful proof (*verunglückter Beweis*).

> While working on the *Grundlagenbuch*, I found myself motivated to rethink Hilbert's second consistency proof for the ε-axiom, the so-called "unsuccessful" proof, and it now seems to me that it can be fixed after all.[53]

This proof bears a striking resemblance to the proof the first ε-theorem in (Hilbert and Bernays, 1939) and to a seven-page sketch in Bernays's hand of a

"consistency proof for the logical axiom of choice" found bound with lecture notes to Hilbert's course on "Elements and principles of mathematics" of 1910.[54] This "unsuccessful" proof seems to me to be another but independent contribution to the development of logic and the ε-calculus, independent of the substituion method. Note that Bernays's proof of Herbrand's theorem in (Hilbert and Bernays, 1939) is based on the (second) ε-theorem is the first correct published proof of that important result.

The realization that the consistency proof even for first-order ε's was problematic came only with Gödel's incompleteness results. In a letter dated March 10, 1931, von Neumann presents an example that shows that in the most recent version of Ackermann's proof, the length of the substitution process not only depends on the rank and degree of ε-terms occurring in the proof, but also on numerical values used as substitutions. He concludes:

> I think that this answers the question, which we recently discussed when going through Ackermann's modified proof, namely whether an estimate of the length of the correction process can be made uniformly and independently of numerical substituends, in the negative. At this point the proof of termination of the procedure (for the next higher degree, i.e., 3) has a gap.[55]

There is no doubt that the discussion of the consistency proof was precipitated by Gödel's results, as both von Neumann and Bernays were aware of these results, and at least von Neumann realized the implications for Hilbert's Program and the prospects of a finitistic consistency proof for arithmetic. Bernays corresponded with Gödel on the relevance of Gödel's result for the viability of the project of consistency proofs just before and after von Neumann's counterexample located the difficulty in Ackermann's proof. On January 18, 1931, Bernays writes to Gödel:

> If one, as does von Neumann, assumes as certain that any finite consideration can be formulated in the framework of System *P*—I think, as you do too, that this is not at all obvious—one arrives at the conclusion that a finite proof of consistency of *P* is impossible.

The puzzle, however, remained unresolved for Bernays even after von Neumann's example, as he writes to Gödel just after the exchange with von Neumann, on April 20, 1931:

> The confusion here is probably connected to that about Ackermann's proof for the consistency of number theory (System $\mathfrak{Z}$), which I have not so far been able to clarify.
>
> That proof—on which Hilbert has reported in his Hamburg talk on the "foundations of mathematics"[56] [...]—I have repeatedly thought through and found correct. On the basis of your results one must now conclude that this proof cannot be formalized within System $\mathfrak{Z}$; indeed, this must hold even if one restricts the system whose consistency is

> to be proved by leaving only addition and multiplication as recursive definitions. On the other hand, I do not see which part of Ackermann's proof makes the formalization within $\mathfrak{Z}$ impossible, in particular if the problem is so restricted.[57]

Gödel's results thus led Bernays, and later Ackermann to reexamine the methods used in the consistency proofs. A completion of the project had to wait until 1940, when Ackermann was able to carry through the termination proof based on transfinite induction—following Gentzen (1936)—on ε_0.

4. Conclusion

With the preceding exposition and analysis of the development of axiomatizations of logic and mathematics and of Hilbert and Ackermann's consistency proofs I hope to have answered some open questions regarding the historical development of Hilbert's Program. Hilbert's *technical* project and its evolution is without doubt of tremendous importance to the history of logic and the foundations of mathematics in the 20th century. Moreover, an understanding of the technical developments can help to inform an understanding of the history and prospects of the *philosophical* project. The lessons drawn in the discussion, in particular, of Ackermann's use of transfinite induction, raise more questions. The fact that transfinite induction in the form used by Ackermann was so readily accepted as finitist, not just by Ackermann himself, but also by Hilbert and Bernays leaves open two possibilities: either they were simply wrong in taking the finitistic nature of Ackermann's proof for granted and the use of transfinite induction simply cannot be reconciled with the finitist standpoint as characterized by Hilbert and Bernays in other writings, or the common view of what Hilbert thought the finitist standpoint to consist in must be revised. Specifically, it seems that the explanation of why transfinite induction is acceptable stresses one aspect of finitism while downplaying another: the *objects* of finitist reasoning are—finite and—intuitively given, whereas the methods of proof were not required to have the epistemic strength that the finitist standpoint is usually thought to require (i.e., to guarantee, in one sense or another, the intuitive evidence of the resulting theorems). Of course, the question of whether Hilbert can make good on his claims that finitistic reasoning affords this intutive evidence of its theorems is one of the main difficulties in a philosphical assessment of the project (see, e.g., Parsons (1998)).

I have already hinted at the implications of a study of the practice of finitism for philosophical reconstructions of the finitist view (in note 4). We are of course free to latch on to this or that aspect of Hilbert's ideas (finitude, intuitive evidence, or surveyability) and develop a philosophical view around it. Such an approach can be very fruitful, and have important and inightful

results (as, e.g, the example of Tait's (1981) work shows). The question is to what extent such a view should be accepted as a reconstruction of Hilbert's view as long as it makes the practice of the technical project come out off base. Surely rational reconstruction is governed by something like a principle of charity. Hilbert and his students, to the extent possible, should be construed so that what they preached is reflected in their practice. This requires, of course, that we know what the practice was. If nothing else, I hope to have provided some of the necessary data for that.

Acknowledgements

Parts of this paper were presented to the Townsend Center Working Group in History and Philosophy of Logic and Mathematics at the University of Caliornia, Berkeley, December 1999, at a workshop on Hilbert at the Institut d'Histoire des Sciences et Techniques of the Université Paris I (partially funded by the Townsend Center through an International Working Group grant), May 2000, at he International Symposium on the History of Logic at the University of Helsinki, June 2000, and at the Department of Logic and Philosophy of Science of the University of California, Irvine, January 2001. I am indebted to Hans Richard Ackermann for providing me with copies of Wilhelm Ackermann's correspondence, and to Wilfried Sieg for providing me with Kneser's notes to the 1921–22 and 1922–23 lectures (Hilbert, 1922a; Hilbert and Bernays, 1923a). I would also like to thank Grigori Mints and Paolo Mancosu for valuable comments. My understanding of the ε-substitution method in general and Ackermann's proof in particular has benefitted greatly from conversations with and writings of Grigori Mints, Georg Moser, and W. W. Tait. Any errors in my presentation are, of course, mine.

Notes

[1] Hilbert (1905, 131). For a general discussion of Hilbert's views around 1905, see Peckhaus (1990, Chapter 3).

[2] „Die von ZERMELO benutze axiomatische Methode ist zwar unanfechtbar und unentbehrlich. Es bleibt dabei doch die Frage offen, ob die aufgestellten Axiome nicht etwa einen Widerspruch einschliessen. Ferner erhebt sich die Frage, ob und inwieweit sich das Axiomensystem aus der Logik ableiten lässt. [... D]er Versuch einer Zurückführung auf die Logik scheint besonders deshalb aussichtsvoll, weil zwischen Mengen, welche ja die Gegenstände in ZERMELOs Axiomatik bilden, und den Prädikaten der Logik ein enger Zusammenhang besteht. Nämlich die Mengen lassen sich auf Prädikate zurückführen.

Diesen Gedanken haben FREGE, RUSSEL[L] und WEYL zum Ausgangspunkt genommen bei ihren Untersuchungen über die Grundlagen der Mathematik." (Hilbert 1920, 27–28).

[3] „Wir müssen uns nämlich fragen, was es bedeuten soll: „es gibt ein Prädikat *P*." In der axiomatischen Mengentheorie bezieht sich das „es gibt" immer auf den zugrunde gelegten

Bereich $\mathfrak{B}$. In der Logik können wir zwar auch die Prädikate zu einem Bereich zusammengefasst denken; aber dieser Bereich der Prädikate kann hier nicht als etwas von vorneherein Gegebenes betrachtet werden, sondern die Prädikate müssen durch logische Operationen gebildet werden, und durch die Regeln der Konstruktion bestimmt sich erst nachträglich der Prädikaten-Bereich.

Hiernach ist ersichtlich, dass bei den Regeln der logischen Konstruktion von Prädikaten die Bezugnahme auf den Prädikaten-Bereich nicht zugelassen werden kann. Denn sonst ergäbe sich ein *circulus vitiosus*.“ (Hilbert 1920, 31).

[4] “RUSSELL geht von dem Gedanken aus, dass es genügt, das zur Definition der Vereinigungsmenge unbrauchbare Prädikat durch ein sachlich gleichbedeutendes zu ersetzen, welches nicht dem gleichen Einwande unterliegt. Allerdings vermag er ein solches Prädikat nicht anzugeben, aber er sieht es als ausgemacht an, dass ein solches existiert. In diesem Sinne stellt er sein „Axiom der Reduzierbarkeit“ auf, welches ungefähr folgendes besagt: „Zu jedem Prädikat, welches durch (ein- oder mehrmalige) Bezugnahme auf den Prädikatenbereich gebildet ist, gibt es ein sachlich gleichbedeutendes Prädikat, welches keine solche Bezugnahme aufweist.

Hiermit kehrt aber RUSSELL von der konstruktiven Logik zu dem axiomatischen Standpunkt zurück. [...]

Das Ziel, die Mengenlehre und damit die gebräuchlichen Methoden der Analysis auf die Logik zurückzuführen, ist heute nicht erreicht und ist vielleicht überhaupt nicht erreichbar.” (Hilbert 1920, 32–33).

[5] “...zur Vermeidung von Paradoxien ist daher eine teilweise gleichzeitige Entwicklung der Gesetze der Logik und der Arithmetik erforderlich.” (Hilbert 1905, 176).

[6] “In Anbetracht der grossen Mannigfaltigkeit von Verknüpfungen und Zusammenhängen, welche die Arithmetik aufweist, ist es von vornherein ersichtlich, dass wir die Aufgabe des Nachweises der Widerspruchslosigkeit nicht mit einem Schlage lösen können. Wir werden vielmehr so vorgehen, dass wir zunächst nur die einfachsten Verknüpfungen betrachten und dann schrittweise immer höhere Operationen und Schlussweisen hinzunehmen, wobei dann für jede Erweiterung des Systems der Zeichen und der Uebergangsformeln einzeln der Nachweis zu erbringen ist, dass sie die auf der vorherigen Stufe festgestellte Widerspruchsfreiheit nicht aufheben.

Ein weiterer wesentlicher Gesichtspunkt ist, dass wir, gemäss unserem Plan der restlosen Formalisierung der Arithmetik, den eigentlich mathematischen Formalismus im Zusammenhang mit dem Formalismus der logischen Operationen entwickeln müssen, sodass—wie ich es ausgedrückt habe—ein simultaner Aufbau von Mathematik und Logik ausgeführt wird.”, (Hilbert 1922b, 8a–9a). The passage is not contained in Kneser’s notes (Hilbert and Bernays, 1923a) to the same course.

[7] The notes by Kneser (Hilbert and Bernays, 1923a) do not contain the list of systems below. The version of the ε-calculus used in the addendum is the same as that used in Kneser’s notes, and differs from the presentation in Ackermann (1924b), submitted February 20, 1924.

[8] (Hilbert and Bernays, 1923b), 17, 19

[9] “**Disposition**. Stufe II war elementares Rechnen Axiome 1–16

Stufe III. Nun elementare Zahlentheorie
Schema für Def. von Funktionen durch Rekursion u. Schlussschema
wollen [?] unser Schlussschema noch das Induktionsschema hinzuziehen
Wenn auch inhaltlich das wesentlich mit den Ergebnissen der anschauliche gewonnenen [?] Zahlenth. übereinstimmt, so doch jetzt Formeln z.B. $a+b=b+a$.

Stufe IIII. Transfinite Schlussweise u. teilweise Analysis

Stufe V. Höhere Variablen-Gattungen u. Mengenlehre. Auswahlaxiom

Stufe VI. Zahlen d[er] 2$^{\text{ten}}$ Zahlkl[asse], Volle transfin[ite] Induktion. Höhere Typen. Continuumsproblem, transfin[ite] Induktion für Zahlen der 2$^{\text{ten}}$ Zahlkl.

Stufe VII. 1.) Ersetzung der ∞ vielen Definitionsschemata durch ein Axiom. 2.) Analysis u[nd] Mengenlehre. Auf der 4$^{\text{ten}}$ Stufe nochmals der volle Satz von der oberen Grenze

Stufe VIII. Formalisierung der Wohlordnung. " (Hilbert and Bernays, 1923a), *Ergänzung*, sheet 1.

[10] The proof can also be found in Hilbert (1922c, 171–173); cf. Mancosu (1998, 208–210).

[11] "Somit sehen wir uns veranlasst, die Beweise als solche zum Gegenstand der Untersuchung zu machen; wir werden zu einer Art von *Beweistheorie* gedrängt, welche mit den Beweisen selbst als Gegenständen operiert.

Für die Denkweise der gewöhnlichen Zahlentheorie sind die Zahlen das gegenständlich-Aufweisbare, und die Beweise der Sätze über die Zahlen fallen schon in das gedankliche Gebiet. Bei unserer Untersuchung ist der Beweis sebst etwas Aufweisbares, und durch das Denken über den Beweis kommen wir zur Lösung unseres Problems.

Wie der Physiker seinen Apparat, der Astronom seinen Standort untersuch, wie der Philosoph Vernunft-Kritik übt, so braucht der Mathematiker diese Beweistheorie, um jeden mathematischen Satz durch eine Beweis-Kritik sicherstellen zu können." Hilbert (1920, 39–40). Almost the same passage is found in Hilbert (1922c, 169–170), cf. Mancosu (1998, 208).

[12] For a detailed discussion of these influences, see Mancosu (1999a).

[13] (Hilbert 1922b, part 2, 3). Kneser's *Mitschrift* of these lectures contains a different system which does not include negation. Instead, numerical inequality is a primitive. This system is also found in Hilbert's first talks on the subject in Copenhagen and Hamburg in Spring and Summer of 1921. Hilbert (1923), a talk given in September 1922, and Kneser's notes to the course of Winter Semester 1922–23 (Hilbert and Bernays, 1923b) do contain the new system with negation. This suggests that the developments of Hilbert's 1921–22 lectures were not incorporated into the published version of Hilbert's Hamburg talk (1922c). Although (1922c) was published in 1922, and a footnote to the title says "This communication is essentially the content of the talks which I have given in the Spring of this year in Copenhagen [. . .] and in the Summer in Hamburg [. . .]," it is clear that the year in question is 1921, when Hilbert addressed the Mathematisches Seminar of the University of Hamburg, July 25–27, 1921. A report of the talks was published by Reidemeister in *Jahrbuch der Deutschen Mathematiker-Vereinigung* **30**, 2. Abt. (1921), 106. Hilbert and Bernays (1923b) also have separate axioms for conjunction and disjunction, while in (1923) it is extended it by quantifiers.

[14] The procedure whereby we pass from $\mathfrak{A}$ to $\mathfrak{A}'$ is simple in this case, provided we keep track of which variables are substituted for below the inference. In general, the problem of deciding whether a formula is a substitution instance of another, and to calculate the substitution which would make the latter syntactically identical to the former is known as *matching*. Although not computationally difficult, it is not entirely trivial either.

[15] "Nennen wir eine Formel, in der keine Variablen und keine Funktionale ausser Zahlzeichen vorkommen, eine „*explizite [numerische] Formel*", so können wir das gefundene Ergebnis so aussprechen: Jede beweisbare explizite [numerische] Formel ist Endformel eines Beweises, dessen sämtliche Formeln explizite Formeln sind.

Dieses müsste insbesondere von der Formel $0 \neq 0$ gelten, wenn sie beweisbar wäre. Der verlangte Nachweis der Widerspruchsfreiheit ist daher erbracht, wenn wir zeigen, dass es keinen Beweis der Formel geben kann, der aus lauter expliziten Formeln besteht.

Um diese Unmöglichkeit einzusehen, genügt es, eine konkret feststellbare Eigenschaft zu finden, die erstens allen den expliziten Formeln zukommt, welche durch Einsetzung aus einem Axiom entstehen, die ferner bei einem Schluss sich von den Prämissen auf die Endformel überträgt, die dagegen nicht auf die Formel $0 \neq 0$ zutrifft." (Hilbert 1922b, part 2, 27–28).

[16] "Wir teilen die expliziten Formeln in „*richtige*" und „*falsche*" ein. Die expliziten Primformeln sind Gleichungen, auf deren beiden Seiten *Zahlzeichen* stehen. Eine solche *Gleichung* nennen wir *richtig*, wenn die beiderseits stehenden Zahlzeichen *übereinstimmen*; andernfalls nennen

wir sie *falsch*. Eine *Ungleichung*, auf deren beiden Seiten Zahlzeichen stehen, nennen wir *richtig*, falls die beiden Zahlzeichen *verschieden* sind; sonst nen[n]en wir sie *falsch*.

In der Normalform einer beliebigen expliziten Formel haben alle Disjunktionsglieder die Gestalt von Gleichungen oder Ungleichungen, auf deren beiden Sieten Zahlzeichen stehen.

Wir nennen nun eine *allgemeine explizite Formel richtig*, wenn in der zugehörigen Normalform jede als Konjunktionsglied auftretende (bezw. die ganze Normalform ausmachende) Disjunktion eine richtige Gleichung oder eine richtige Ungleichung als Glied enthält. Andernfalls nennen wir die Formel *falsch*. [...]

Nach der gegebenen Definition lässt sich die Frage, ob eine explizierte [sic] Formel richtig oder falsch ist, in jed[e]m Falle *konkret entscheiden*. Hier gilt also das „tertium non datur" [...]" (Hilbert 1922b, part 2, 33).

[17] Hilbert (1918b, 149-150). See also Zach (1999, §2.3).

[18] A sketch of the consistency proof is found in the Kneser *Mitschrift* to the 1921–22 lectures (Hilbert, 1922a) in Heft II, pp. 23–32 and in the official notes by Bernays (Hilbert 1922b, part 2, 19–38). The earlier Kneser *Mitschrift* leaves out step (1), and instead of eliminating variables introduces the notion of *einsetzungsrichtig* (correctness under substitution, i.e., every substitution instance is correct). These problems were avoided in the official Bernays typescript. The Kneser notes did contain a discussion of recursive definition and induction, which is not included in the official notes; more about these in the next section.

[19] In the 1921–22 lectures, it is initially argued that the result of applying transformations (1)–(3) results in a *proof* of the same end formula (if substitutions are added to the initial formulas). Specifically, it is suggested that the result of applying elimination of variables and reduction of functionals to the axioms results in formulas which are substitution instances of axioms. It was quickly realized that this is not the case. (When Bernays presented the proof in the 1922–23 lectures on December 14, 1922, he comments that the result of the transformation need not be a proof (Hilbert and Bernays 1923b, 21). The problem is the axiom of equality

$$a = b \to (A(a) \to A(b)).$$

Taking $A(c)$ to be $\delta(c) = c$, a substitution instance would be

$$0+1+1 = 0 \to (\delta(0+1+1) = 0+1+1 \to \delta(0) = 0)$$

This reduces to

$$0+1+1 = 0 \to (0+1 = 0+1+1 \to 0 = 0)$$

which is not a substitution instance of the equality axiom. The consistency proof itself is not affected by this, since the resulting formula is still correct (in Hilbert's technical sense of the word). The official notes to the 1921–22 lectures contain a 2-page correction in Bernays's hand (Hilbert 1922b, part 2, between pp. 26 and 27).

[20] The induction rule is not used in (Ackermann, 1924b), since he deals with stage III only in passing and attempts a consistency proof for all of analysis. There, the induction rule is superseded by an ε-based induction axiom. For a consistency proof of stage III alone, an induction rule is needed, since an axiom cannot be formulated without quantifiers (or ε). The induction rule was introduced for stage III in the Kneser notes to the 1921–22 lectures (Hilbert, 1922a, Heft II, 32) and the 1922–23 lectures (Hilbert and Bernays, 1923b, 26). It is not discussed in the official notes or the publications from the same period (Hilbert, 1922c; 1923

[21] The general tenor, outlook, and aims of Skolem's work are sufficiently different from Hilbert's to suggest there was no influence either way. Skolem states in his concluding remarks that he wrote the paper in 1919, after reading Russell and Whitehead's *Principia Mathematica*. However, neither Hilbert nor Bernays's papers contain an offprint or manuscript of Skolem's paper, nor correspondence. Skolem is not cited in any of Hilbert's, Bernays's, or Ackermann's papers of the period, although the paper is referenced in (Hilbert and Bernays, 1934).

[22] "Uns fehlt noch ganz das Ax[iom] der vollst[ändigen] Induktion. Man könnte meinen, es wäre

$$\{Z(a) \to (A(a) \to A(a+1))\} \to \{A(1) \to (Z(b) \to A(b))\}$$

Das ist es nicht; denn man setze $a = 1$. Die Voraussetzung muß für *alle a* gelten. Wir haben aber noch gar kein Mittel, das *Alle* in die Voraussetzung zu bringen. Unser Formalismus reich noch nicht hin, das Ind.ax. aufzuschreiben.

Aber als Schema können wir es: Wir erweitern unsere Beweismethoden durch das nebenstehende Schema

$$\frac{\mathfrak{K}(1) \quad \mathfrak{K}(a) \to \mathfrak{K}(a+1)}{Z(a) \to \mathfrak{K}(a)}$$

Jetzt ist es vernünftig, zu fragen, ob dies Schema zum Wspruch führen kann." (Hilbert 1922a, 32). Z is the predicate expressing "is a natural number," it disappears from later formulations of the schema.

[23] "Wie ist es bei Rekursionen? $\varphi(\mathfrak{z})$ komme vor. Entweder 0, dann setzten wir $\mathfrak{a}$ dafür. Oder $\varphi(\mathfrak{z}+1)$: $\mathfrak{b}(\mathfrak{z}, \varphi\mathfrak{z})$. Beh[auptung]: Das Einsetzen kommt zu einem Abschluß, wenn wir zu innerst anfangen." (Hilbert and Bernays 1923a, 29)

[24] "Nicht endlich (durch Rek[ursion]) definiert ist z.B. $\varphi(a) = 0$ wenn es ein b gibt, so daß $a^5 + ab^3 + 7$ Primz[ahl] ist sonst $= 1$. Aber erst bei solchen Zahlen und Funktionen beginnt das eigentliche math[ematische] Interesse, weil dort die Lösbarkeit in endlich vielen Schritten nicht vorauszusehen ist. Wir haben die Überzeugung, daß solche Fragen wie nach dem Wert $\varphi(a)$ lösbar, d.h. daß $\varphi(a)$ doch endlich definierbar ist. Darauf können wir aber nicht warten: wir müssen solche Definitionen zulassen, sonst würden wir den freien Betrieb der Wissenschaft einschränken. Auch den Begriff der Funktionenfunktion brauchen wir." (Hilbert 1922a, Heft III, 1–2).

[25] A full proof is given by Ackermann (1924b).

[26] "Als erstes zeigt man, daß man alle Variablen fortschaffen kann, weil auch hier nur freie Var[iable] vorkommen. Wir suchen die innersten τ und α. Unter diesen stehen nur endlich definierte Funkt[ionen] φ, φ'... Unter diesen können einige im Laufe des Beweises für f in die Ax[iome] eingesetzt sein. 1: $\tau(\varphi) = 0 \to (Z(\mathfrak{a} \to \varphi\mathfrak{a} = 1)$ wo $\mathfrak{a}$ ein Funktional ist. Wenn dies *nicht* benutzt wird, setze ich alle $\alpha(\varphi)$ und $\tau(\varphi)$ gleich Null. Sonst reduziere ich $\mathfrak{a}$ und $\varphi(\mathfrak{a})$ und sehe, ob $Z(\mathfrak{a} \to \varphi(\mathfrak{a}) = 1$ in allen ... wo sie vorkommt, richtig ist. Ist die richtig, so setze ich $\tau = 0$ $\alpha = 0$. Ist sie falsch, d.h. is $\mathfrak{a} = \mathfrak{z}$ $\varphi(\mathfrak{z} \neq 1$, so setzen wir $\tau(\varphi) = 1$, $\alpha(\varphi) = \mathfrak{z}$. Dabei bleibt der Beweis Beweis. Die an Stelle der Axiome gesetzten Formeln sind richtig.

Der Gedanke ist: wenn ein Beweis vorliegt, so kann ich aus ihm ein Argument finden für das $\varphi = 1$ ist). So beseitigt man schrittweise die τ und α aund Anwendungen von 1 2 3 4 und erhält einen Beweis von $1 \neq 1$ aus I–V und richtigen Formeln d.h. aus I–V,

$$\begin{aligned}
\tau(f,b) = 0 &\to \{Z(a) \to f(a,b) = 1\} \\
\tau(f,b) \neq 0 &\to Z(f(\alpha,b)) \\
\tau(f,b) \neq 0 &\to f(\alpha(f,b),b) \neq 1 \\
\tau(f,b) \neq 0 &\to \tau(f,b) = 1\text{"}
\end{aligned}$$

(Hilbert, 1922a), Heft III, 3–4. The lecture is dated February 23, 1922.

[27] "Was fehlt uns?

1. in logischer Hinsicht. Wir haben nur gehabt den Aussagenkalkül mit der Erweiterung auf freie Variable d.h. solche für die beliebige Funktionale eingesetzt werden konnten. Es fehlt das Operieren mit „alle" und „es gibt".

2. Wir haben das Induktionsschema hinzugefügt, ohne W[iderspruchs]-f[reiheits] Beweis und auch nur provisorisch, also in der Absicht, es wegzuschaffen.

3. Bisher nur die arithmethischen Axiome genau [?] die sich auf ganze Zahlen beziehen. Und die obigen Mängel verhindern uns ja natürlich die Analysis aufzubauen (Grenzbegriff, Irrationalzahl).

Diese 3 Punkte liefern schon Disposition und Ziele für das Folgende.

Wir wenden uns zu 1.) Es ist ja an sich klar, dass eine Logik ohne „alle"—„es gibt" Stückwerk wäre, ich erinnere wie gerade in der Anwendung dieser Begriffe, und den sogennanten transfiniten Schlussweisen die Hauptschwierigkeiten entstanden. Die Frage der Anwendbarkeit dieser Begriffe auf ∞ Gesamtheiten haben wir noch nicht behandelt. Nun könnten wir so verfahren, wie wir es beim Aussagen-Kalkül gemacht haben: einige, möglichst einfache [???] als Axiome zu formalisieren, aus denen sich [sic] dann alle übrigen folgen. Dann müsste der W-f Beweis geführt werden—unserem allgem[einen] Programm gemäss: mit unserer Einstellung, dass Beweis eine vorliegende Figur ist. Für den W-f Beweis grosse Schwierigkeiten wegen der gebundenen Variabeln. Die tiefere Untersuchung zeigt aber, dass der eigentliche Kern der Schwierigkeit an einer anderen Stelle liegt, auf die man gewöhnlich erst später Acht giebt und die auch in der Litteratur erst später wahrgenommen worden ist."(Hilbert and Bernays 1923b, *Ergänzung*, sheets 3–4).

[28] "[Dieser Kern liegt] beim *Auswahlaxiom* von Zermelo. [...] Die Einwände richten sich gegen das Auswahlprinzip. Sie müßten sich aber ebenso gegen „alle" und „es gibt" richten, wobei derselbe Grundgedanke zugrunde liegt.

Wir wollen das Auswahlaxiom erweitern. Jeder Aussage mit einer Variablen $A(a)$ ordnen wir ein Ding zu, für das die Aussage nur dann gilt, wenn sie allgemein gilt. Also ein Gegenbeispiel, wenn es existiert.

$\varepsilon(A)$, eine individuelle logische Funktion. [...] ε genüge dem *transfiniten Axiom*:

(16) $\quad A(\varepsilon A) \to Aa$

z.B. Aa heiße: a ist bestechlich. εA ist Aristides."(Hilbert and Bernays 1923a, 30–31). The lecture is dated February 1, 1922, given by Hilbert. The corresponding part of Hilbert's notes for that lecture in (Hilbert and Bernays 1923b, Ergänzung, sheet 4)) contains page references to (Hilbert 1923, 152 and 156, paras. 4–6 and 17–19 of the English translation), and indicates the changes made for the lecture, specifically, to replace τ by ε.

[29] See section 3.4 on the ε-substitution method.

[30] "Wenn wir eine *Funktionsvariable* haben:

$$A\varepsilon_f Af \to Af$$

(π fällt fort)? ε komme *nur* mit $\mathfrak{A}$ vor (z.B. $f0 = 0$, $ff0 = 0$). Wie werden wir die Funktionsvariablen ausschalten? Statt fc setzen wir einfach c. Auf die *gebundenen* trifft das nicht zu. Für diese nehmen wir probeweise eine bestimmte Funktion z.B. δa und führen damit die Reduktion durch. Dann steht z.B. $\mathfrak{A}\delta \to \mathfrak{A}\varphi$. Diese reduziert ist r[ichtig] oder f[alsch]. Im letzten Falle is $\mathfrak{A}\varphi$ falsch. Dann setzen wir überall φ für $\varepsilon_f \mathfrak{A} f$. Dann steht $\mathfrak{A}\varphi \to \mathfrak{A}\psi$. Das ist sicher *r[ichtig]* da $\mathfrak{A}\varphi$ f[alsch] ist." (Hilbert and Bernays 1923a, 38–39).

[31] For a more detailed survey of Ackermann's scientific contributions, see (Hermes, 1967). A very informative discussion of Ackermann's scientific correspondence can be found in (Ackermann, 1983).

[32] "In seiner Arbeit "Begründung des „Tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit hat Ackermann im allgemeinsten Falle gezeigt, dass der Gebrauch der Worte „alle" und „es gibt", des „Tertium non datur" widerspruchsfrei ist. Der Beweis erfolgt unter ausschliesslicher Benutzung primitiver und endlicher Schlussweisen. Es wird alles an dem mathematischen Formalismus sozusagen direkt demonstriert.

Ackermann hat damit unter Ueberwindung erheblicher mathematischer Schwierigkeiten ein Problem gelöst, das bei den modernene auf eine Neubegründung der Mathematik gerichteten

Bestrebungen an erster Stelle steht." Hilbert-Nachlaß, Niedersächsische Staats- und Universitätsbibliothek Göttingen, Cod. Ms. Hilbert 458, sheet 6, no date. The three-page letter was evidently written in response to a request by the President of the International Education Board, dated May 1, 1924.

[33] "Ich bemerke nur, dass Ackermann meine Vorlesungen über die Grundlagen der Math-[ematik] in den letzten Semestern gehört hat und augenblicklich einer der besten Herren der Theorie ist, die ich hier entwickelt habe." *ibid.*, sheet 2. The draft is dated March 19, 1924, and does not mention Russell by name. Sieg (1999), however, quotes a letter from Russell's wife to Hilbert dated May 20, 1924, which responds to an inquiry by Hilbert concerning Ackermann's stay in Cambridge. Later in the letter, Hilbert expresses his regret that the addressee still has not been able to visit Göttingen. Sieg documents Hilbert's effort in the preceding years to effect a meeting in Göttingen; it is therefore quite likely that the addressee was Russell.

[34] Ackermann only requires that b be bound by the occurrence of φ, but this is not enough for his proof.

[35] According to Ackermann's definition of subordination, this would not be true. A subterm of $\mathfrak{c}(b)$ might contain a bound variable and thus not be a constant subterm, but the variable could be bound by a function symbol in $\mathfrak{t}$ other than the occurrence of φ under consideration. See note 4.

[36] Ackermann (1924b, 18)

[37] Tait (1981) argues that finitism coincides with primitive recursive arithmetic, and that therefore the Ackermann function is not finitistic. Tait does not present this as a historical thesis, and his conceptual analysis remains unaffected by the piece of historical evidence presented here. For further evidence (dating however mostly from after 1931) see Zach (1998, §5) and Tait's response in (2000).

[38] Ackermann (1924b, 13–14).

[39] "[Gentzen fragt,] ob Sie der Meinung sind, dass sich die Methode des Endlichkeitsbeweises durch transfinite Induktion auf den Wf-Beweis Ihrer Dissertation anwenden lasse. Ich würde es sehr begrüssen, wenn das ginge." Bernays to Ackermann, November 27, 1936, Bernays Papers, ETH Zürich Library/WHS, Hs 975.100.

[40] "Mir fällt übrigens jetzt, wo ich gerade meine Dissertation zur Hand nehme, auf, dass dort in ganz ähnlicher Weise mit transfiniten Ordnungszahlen operiert wird wie bei Gentzen." Ackermann to Bernays, December 5, 1936, Bernays Papers, ETH Zürich Library/WHS, Hs 975.101.

[41] "Ich weiss übrigens nicht, ob Ihnen bekannt ist (ich hatte das seiner Zeit nicht als Ueberschreitung des engeren finiten Standpunktes empfunden), dass in meiner Dissertation transfinite Schlüsse benutzt werden. (Vgl. z.B. die Bemerkungen letzter Abschnitt Seite 13 und im nächstfolgenden Abschnitt meiner Dissertation." Ackermann to Bernays, June 29, 1938, Bernays papers, ETH-Zürich, Hs 975.114. The passage Ackermann refers to is the one quoted above.

[42] The ε-substitution method was subsequently refined by von Neumann (1927) and Hilbert and Bernays (1939). Ackermann (1940) gives a consistency proof for first-order arithmetic, using ideas of Gentzen (1936). See also (Tait, 1965) and (Mints, 1994). Useful introductions to the ε-substitution method of Ackermann (1940) and to the ε-notation in general can be found in Moser (2000) and Leisenring (1969), respectively.

[43] Ackermann (1924b, 8). The π-functions were present in the earliest presentations in (Hilbert, 1922a) as the τ-function and also occur in (Hilbert and Bernays, 1923a). They were dropped from later presentations.

[44] It is not clear whether the definition is supposed to apply to the formulas with free variables (i.e., to $a = b$ and $a = \varepsilon_b(a = b)$ in the example) or to the corresponding substituion instances. The proof following the definition on p. 21 of (Ackermann, 1924b) suggests the former, however, later in the procedure for defining a sequence of total substitutions it

is suggested that the ε-expressions corresponding to formulas subordinate to $\tilde{\mathfrak{A}}(a)$ receive substitutions—but according to the definition of a total substitution only ε-*terms* ($\varepsilon_b(\mathfrak{z} = b)$ in the example) receive substitutions.

[45] The bookkeeping functions are introduced here and are not used by Ackermann. The basic idea is that that in case (3), substitutions for some formulas are discarded, and the next substitution is given the "last" total substitution where the substitution for the formula was not yet marked as discarded. Instead of explicit bookkeeping, Ackermann uses the notion of a formula being "remembered" as having its value not discarded.

[46] With the restriction on second-order ε-terms imposed by Ackermann, and discussed below, the system for which a consistency proof was claimed is essentially elementary analysis, a predicative system. A consistency proof using the ε-substitution method for this system was given by Mints and Tupailo (1996).

[47] Ackermann (1924b, 9)

[48] Ackermann to Bernays, June 25, 1925, Bernays Papers, ETH Zürich, Hs. 975.96.

[49] "Ich habe augenblicklich den ε_f-Beweis wieder vorgenommen, und versuche mit aller Gewalt da zum Abschluß zu kommen. Daß sich das Problem auf ein zahlentheoretisches reduzieren läßt, hatte ich Ihnen damals ja schon mitgeteilt. Den zahlentheoretischen Satz allgemein zu beweisen scheint mir aber ebenso schwierig wie das ganze Problem. Ich habe nun den schon mehrfach von mir versuchten Weg wieder eingeschlagen, den Begriff des Grundtyps so zu erweitern, das auch die ε mit freien Funktionsvariablen eine Ersetzung bekommen. Dieser Weg scheint ja auch der natürlichste, und die Gleichheitsaxiome $(f)(A(f) \rightleftharpoons B(f)) \to \varepsilon_f A f \equiv \varepsilon_f B f$ würden dann gleich mitbehandelt. Ich habe einige Hoffnung, daß die sich früher auf diesem Weg einstellenden Schwierigkeiten vermieden werden können, wenn ich den ε_a-Formalismus benutze und statt ohne ε definierte Funktionen, solche zur Ersetzung für die ε_f nehme, die ein ε_a enthalten können. Ich habe mir aber erst einfache Spezialfälle überlegt." Ackermann to Bernays, March 31, 1926. ETH Zürich/WHS, Hs 975.97. Although Ackermann's mention of "ground types" precedes the publication of von Neumann (1927), the latter paper was submitted for publication alread on July 29, 1925.

[50] "Letzthin habe ich mir Ihren neueren Beweis der Widerspruchsfr[eiheit] für die ε_a an Hand dessen, was Sie mir vor Ihrer Abreise aufschrieben, genauer überlegt und glaube diesen Beweis als richtig eingesehen zu haben." Bernays to Ackermann, April 12, 1927, in the posession of Hans Richard Ackermann. Bernays continues to remark on specifics of the proof, roughly, that when example substitutions for ε-types are revised (the situation corresponding to case (3) in Ackermann's original proof), the substitutions for types of higher rank have to be reset to the initial substitution. He gives an example that shows that if this is not done, the procedure does not terminate. He also suggests that it would be more elegant to treat all types of the same rank at the same time and gives an improved estimate for the number of steps necessary. Note that the reference to "ε_a's" (as opposed to ε_f) suggest that the proof was only for the first-order case. A brief sketch of the proof is also contained in a letter from Bernays to Weyl, dated January 5, 1928 (ETH Zürich/WHS, Hs. 91.10a).

[51] "Wie Sie sich vielleicht erinnern, hatte ich damals einen 2. Beweis für die Widerspruchsfreiheit der ε_a. Dieser Beweis ist von mir nie publiziert worden, sondern nur Herrn Prof. Bernays mündlich mitgeteilt worden, der sich auch damals von seiner Richtigkeit überzeugte. Prof. Bernays schrieb mir nun im vergangenen Jahre, daß das Ergebnis ihm mit der Gödelschen Arbeit nicht zu harmonisieren scheine." Ackermann to Hilbert, August 23, 1933, Hilbert-Nachlaß, Niedersächsische Staats- und Universitätsbibliothek, Cod. Ms. Hilbert 1. Ackermann did not then locate the difficulty, and even a year and a half later (Ackermann to Bernays, December 8, 1934, ETH Zürich/WHS, Hs 975.98) suggested a way that a finitistic consistency proof of arithmetic could be found based on work of Herbrand and Bernays's drafts for the second volume of *Grundlagen*.

[52] "Problem I. The consistency proof of the ε-axiom for the function variable f. We have the outline of a proof. Ackermann has already carried it out to the extent that the only remaining task consists in the proof of an elementary finiteness theorem that is purely arithmetical." Hilbert (1928), translated in Mancosu (1999a, 229). The extension to ε-extensionality is Problem III.

[53] "Anlässlich der Arbeit für das Grundlagenbuch sah ich mich dazu angetrieben, den zweiten Hilbertschen Wf.-Beweis für das ε-Axiom, den sogenannten „verunglückten" Beweis, nochmals zu überlegen, und es scheint mir jetzt, dass dieser sich doch richtig stellen lässt." Bernays to Ackermann, October 16, 1929, in the possession of Hans Richard Ackermann. Bernays continues with a detailed exposition of the proof, but concludes that the proof probably cannot be extended to include induction, for which ε-substitution seems better suited.

[54] The sketch bears the title "Wf.-Beweis für das logische Auswahl-Axiom", and is inserted in the front of *Elemente und Prinzipienfragen der Mathematik*, Sommer-Semester 1910. Library of the Mathematisches Institut, Universität Göttingen, 16.206t14. A note in Hilbert's hand says "Einlage in W.S. 1920." However, the ε-Axiom used is the more recent version $Ab \to A\varepsilon_a A_a$ and not the original, dual $A\varepsilon_a Aa \to Ab$. It is thus very likely that the sketch dates from after 1923.

[55] "Ich glaube, dass damit die Frage, die wir bei der Durchsprechung des modifizierten Ackermannschen Beweises zuletzt diskutierten, ob nämlich eine Längen-Abschätzung für das Korrigier-Verfahren unabhängig von der Grösse der Zahlen-Substituenden gleichmässig möglich sei, verneinend beantwortet ist. An diesem Punkte ist dann der Nachweis des endlichen Abbrechens dieses Verfahrens (für den nächsten Grad, d.h. 3) jedenfalls lückenhaft.)" von Neumann to Bernays, March 10, 1931, Bernays Papers, ETH Zürich/WHS, Hs. 975.3328. Von Neumann's example can be found in Hilbert and Bernays (1939, 123).

[56] Hilbert (1928).

[57] In a letter dated May 3, 1931, Bernays suggests that the problem lies with certain types of recursive definitions The Bernays–Gödel correspondence will shortly be published in Volume IV of Gödel's collected works. For more on the reception of Gödel's results by Bernays and von Neumann, see Dawson (1988) and Mancosu (1999b).

References

Ackermann, H. R.: 1983, 'Aus dem Briefwechsel Wilhelm Ackermanns'. *History and Philosophy of Logic* **4**, 181–202.

Ackermann, W.: 1924a, 'Begründung des "tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit'. Dissertation, Universität Göttingen.

Ackermann, W.: 1924b, 'Begründung des "tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit'. *Mathematische Annalen* **93**, 1–36.

Ackermann, W.: 1928a, 'Über die Erfüllbarkeit gewisser Zählausdrücke'. *Mathematische Annalen* **100**, 638–649.

Ackermann, W.: 1928b, 'Zum Hilbertschen Aufbau der reellen Zahlen'. *Mathematische Annalen* **99**, 118–133.

Ackermann, W.: 1940, 'Zur Widerspruchsfreiheit der Zahlentheorie'. *Mathematische Annalen* **117**, 162–194.

Ackermann, W.: 1954, *Solvable Cases of the Decision Problem*. Amsterdam: North-Holland.

Dawson, J. W.: 1988, 'The Reception of Gödel's Incompleteness Theorems'. In: S. G. Shanker (ed.): *Gödel's Theorem in Focus*. London and New York: Routledge, pp. 74–95.

Dedekind, R.: 1888, *Was sind und was sollen die Zahlen?* Braunschweig: Vieweg. English translation in (Dedekind, 1901).

Dedekind, R.: 1901, *Essays on the Theory of Number*. New York: Open Court.

Ewald, W. B. (ed.): 1996, *From Kant to Hilbert. A Source Book in the Foundations of Mathematics*, Vol. 2. Oxford: Oxford University Press.

Gentzen, G.: 1936, 'Die Widerspruchsfreiheit der reinen Zahlentheorie'. *Mathematische Annalen* **112**, 493–565.

Hermes, H.: 1967, 'In memoriam Wilhelm Ackermann, 1896–1962'. *Notre Dame Journal of Formal Logic* **8**, 1–8.

Hilbert, D.: 1900, 'Mathematische Probleme'. *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Math.-Phys. Klasse* pp. 253–297. Lecture given at the International Congress of Mathematicians, Paris, 1900. Partial English translation in Ewald (1996, 1096–1105).

Hilbert, D.: 1905, 'Über die Grundlagen der Logik und der Arithmetik'. In: A. Krazer (ed.): *Verhandlungen des dritten Internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*. Leipzig, pp. 174–85. English translation in van Heijenoort (1967, 129–38).

Hilbert, D.: 1918a, 'Axiomatisches Denken'. *Mathematische Annalen* **78**, 405–15. Lecture given at the Swiss Society of Mathematicians, 11 September 1917. Reprinted in Hilbert (1935, 146–56). English translation in Ewald (1996, 1105–1115).

Hilbert, D.: 1918b, 'Prinzipien der Mathematik'. Lecture notes by Paul Bernays. Winter-Semester 1917–18. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, D.: 1920, 'Probleme der mathematischen Logik'. Vorlesung, Sommer-Semester 1920. Lecture notes by Paul Bernays and Moses Schönfinkel. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, D.: 1922a, 'Grundlagen der Mathematik'. Vorlesung, Winter-Semester 1921–22. Lecture notes by Helmut Kneser. Unpublished manuscript, three notebooks.

Hilbert, D.: 1922b, 'Grundlagen der Mathematik'. Vorlesung, Winter-Semester 1921–22. Lecture notes by Paul Bernays. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, D.: 1922c, 'Neubegründung der Mathematik: Erste Mitteilung'. *Abhandlungen aus dem Seminar der Hamburgischen Universität* **1**, 157–77. Reprinted with notes by Bernays in Hilbert (1935, 157–177). English translation in Ewald (1996, 1115–1134).

Hilbert, D.: 1923, 'Die logischen Grundlagen der Mathematik'. *Mathematische Annalen* **88**, 151–165. Lecture given at the Deutsche Naturforscher-Gesellschaft, September 1922. Reprinted in Hilbert (1935, 178–191). English translation in Ewald (1996, 1134–1148).

Hilbert, D.: 1926, 'Über das Unendliche'. *Mathematische Annalen* **95**, 161–90. Lecture given Münster, 4 June 1925 . English translation in van Heijenoort (1967, 367–392).

Hilbert, D.: 1928, 'Die Grundlagen der Mathematik'. *Abhandlungen aus dem Seminar der Hamburgischen Universität* **6**, 65–85. English translation in van Heijenoort (1967, 464-479).

Hilbert, D.: 1935, *Gesammelte Abhandlungen*, Vol. 3. Berlin: Springer.

Hilbert, D. and W. Ackermann: 1928, *Grundzüge der theoretischen Logik*. Berlin: Springer.

Hilbert, D. and P. Bernays: 1923a, 'Logische Grundlagen der Mathematik'. Winter-Semester 1922–23. Lecture notes by Helmut Kneser. Unpublished manuscript.

Hilbert, D. and P. Bernays: 1923b, 'Logische Grundlagen der Mathematik'. Vorlesung, Winter-Semester 1922–23. Lecture notes by Paul Bernays, with handwritten notes by Hilbert. Hilbert-Nachlaß, Niedersächsische Staats- und Universitätsbibliothek, Cod. Ms. Hilbert 567.

Hilbert, D. and P. Bernays: 1934, *Grundlagen der Mathematik*, Vol. 1. Berlin: Springer.

Hilbert, D. and P. Bernays: 1939, *Grundlagen der Mathematik*, Vol. 2. Berlin: Springer.

Leisenring, A. C.: 1969, *Mathematical Logic and Hilbert's ε-Symbol*. London: MacDonald.

Mancosu, P. (ed.): 1998, *From Brouwer to Hilbert. The Debate on the Foundations of Mathematics in the 1920s*. Oxford: Oxford University Press.

Mancosu, P.: 1999a, 'Between Russell and Hilbert: Behmann on the foundations of mathematics'. *Bulletin of Symbolic Logic* **5**(3), 303–330.

Mancosu, P.: 1999b, 'Between Vienna and Berlin: The Immediate Reception of Gödel's Incompleteness Theorems'. *History and Philosophy of Logic* **20**, 33–45.

Mints, G. and S. Tupailo: 1996, 'Epsilon Substitution Method for Elementary Analysis'. *Archive for Mathematical Logic* **35**, 103–130.

Mints, G. E.: 1994, 'Gentzen-type Systems and Hilbert's Epsilon Substitution Method. I'. In: D. Prawitz, B. Skyrms, and D. Westerståhl (eds.): *Logic, Methodology and Philosophy of Science IX*. Amsterdam: Elsevier, pp. 91–122.

Moser, G.: 2000, 'The Epsilon Substitution Method'. Master's thesis, University of Leeds.

Parsons, C.: 1998, 'Finitism and intuitive knowledge'. In: M. Schirn (ed.): *The Philosophy of Mathematics Today*. Oxford: Oxford University Press, pp. 249–270.

Peckhaus, V.: 1990, *Hilbertprogramm und Kritische Philosophie*. Göttingen: Vandenhoeck und Ruprecht.

Sieg, W.: 1999, 'Hilbert's Programs: 1917–1922'. *Bulletin of Symbolic Logic* **5**(1), 1–44.

Skolem, T.: 1923, 'Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich'. *Videnskapsselskapets skrifter I. Matematisk-naturvidenskabelig klasse* **6**. English translation in van Heijenoort (1967, 302–333).

Tait, W. W.: 1965, 'The Substitution Method'. *Journal of Symbolic Logic* **30**, 175–192.

Tait, W. W.: 1981, 'Finitism'. *Journal of Philosophy* **78**, 524–546.

Tait, W. W.: 2000, 'Remarks on Finitism'. In: W. Sieg, R. Sommer, and C. Talcott (eds.): *Reflections. A Festschrift Honoring Solomon Feferman*. forthcoming.

van Heijenoort, J. (ed.): 1967, *From Frege to Gödel. A Source Book in Mathematical Logic, 1897–1931*. Cambridge, Mass.: Harvard University Press.

von Neumann, J.: 1927, 'Zur Hilbertschen Beweistheorie'. *Mathematische Zeitschrift* **26**, 1–46.

Zach, R.: 1998, 'Numbers and Functions in Hilbert's Finitism'. *Taiwanese Journal for Philosophy and History of Science* **10**, 33–60.

Zach, R.: 1999, 'Completeness before Post: Bernays, Hilbert, and the development of propositional logic'. *Bulletin of Symbolic Logic* **5**(3), 331–366.

Address for Offprints: Group in Logic and the Methodology of Science
University of California, Berkeley
Berkeley, CA 94720–3840
http://www.math.berkeley.edu/∼zach/